Universidade de Aveiro Departamento de
Ano 2014 Electrónica, Telecomunicações e Informática

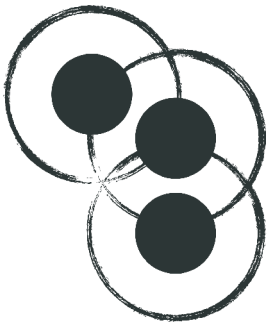
Sérgio Ivan Fernandes Lopes

In search of reliable centimeter-level indoor positioning. A smartphone-based approach.

Desenvolvimento de metodologias para localização *indoor* de *smartphones* com exatidão ao centímetro.

Akkurate™

Smartphone Indoor Positioning For Pervasive Applications



Sérgio Ivan Fernandes Lopes

In search of reliable centimeter-level indoor positioning. A smartphone-based approach.

Tese apresentada à Universidade de Aveiro para cumprimento dos requisitos necessários à obtenção do grau de Doutor em Engenharia Eletrotécnica, realizada sob a orientação científica do Doutor José Manuel Neto Vieira, Professor Auxiliar do Departamento de Eletrónica, Telecomunicações e Informática da Universidade de Aveiro e do Doutor Nuno Miguel Gonçalves Borges de Carvalho, Professor Catedrático do Departamento de Eletrónica, Telecomunicações e Informática da Universidade de Aveiro.

To my wife Susana and my child João
without whom this thesis work
would have been rather dull.

o júri / the jury

presidente / president

Professor Doutor Helmuth Robert Malonek
Professor Catedrático da Universidade de Aveiro

vogais / examiners committee

Professor Doutor Jesús Ureña Ureña
Professor Catedrático da Universidade de Alcalá, Madrid, Espanha

Professor Doutor Adriano Jorge Cardoso Moreira
Professor Associado da Universidade do Minho

Professor Doutor Nuno Miguel da Costa Santos Fonseca
Professor Adjunto do Instituto Politécnico de Leiria

Professor Doutor António Guilherme Rocha Campos
Professor Auxiliar da Universidade de Aveiro

Professor Doutor José Manuel Neto Vieira
Professor Auxiliar da Universidade de Aveiro

agradecimentos /
acknowledgements

O meu primeiro agradecimento vai para os meus orientadores Professor José Manuel Neto Vieira e Professor Nuno Miguel Gonçalves Borges de Carvalho, pelo apoio, orientação, e partilha de conhecimentos. Um agradecimento adicional ao Professor José Manuel Neto Vieira pela motivação constante e elevada disponibilidade. Gostaria também de agradecer aos professores Guilherme Campos e Paulo Dias pelas críticas que foram fazendo e que de algum modo contribuíram para o desenvolvimento do trabalho aqui apresentado.

Agradeço também ao Professor Luís Jesus pela disponibilização da câmara insonorizada do SLHlab/ESSUA para os testes de avaliação perceptual com utilizadores, e ao Daniel Pape pela ajuda e apoio na execução dos mesmos.

Deixo aqui também um agradecimento especial ao Daniel Albuquerque pelas críticas e partilha de ideias, que acrescentaram valor ao trabalho desenvolvido.

Por fim, agradeço aos meus pais o sacrifício que sempre fizeram para me oferecer as melhores condições para estudar.

palavras-chave

Localização de *Smartphones*, Localização com exatidão ao centímetro, Áudio não-invasivo, Redes de Sensores Sem Fios (RSSF), Sincronização de Relógios.

resumo

Esta tese descreve o projeto e a implementação de um sistema de localização para ambientes interiores totalmente compatível com um *smartphone* convencional. O sistema proposto explora a capacidade de aquisição de sinais áudio e de processamento do *smartphone* para medir distâncias utilizando sinais acústicos na banda do audível; foram utilizados sinais áudio não-invasivos, i.e. com reduzido impacto perceptual em humanos. No desenvolvimento deste sistema foram consideradas aplicações que exigem elevada exatidão, na ordem dos centímetros, tais como realidade aumentada, realidade virtual, jogos ou guias virtuais. Utilizou-se uma infraestrutura de faróis de baixo custo suportada por uma rede de sensores sem fios (RSSF). Para manter a infraestrutura síncrona, foi desenvolvido um protocolo de sincronização e sintonização automática, (*Automatic Time Synchronization and Syntonization* - ATSS) que garante um desvio padrão do erro de *offset* abaixo de 1.25 μs. Cada *smartphone* efectua medidas MT-TDoA que posteriormente são utilizadas pelo algoritmo de localização hiperbólica. As estimativas de posição resultantes são estáveis e precisas, com um erro médio absoluto menor do que 10 cm em 95% dos casos e um desvio padrão médio de 2.2 cm, para um período de atualização de posição de 350 ms.

keywords

Smartphone Positioning, Centimeter-level Positioning, Non-Invasive Audio, Wireless Sensor Networks (WSN), Clock Synchronization.

abstract

This thesis describes the design and implementation of a reliable centimeter-level indoor positioning system fully compatible with a conventional smartphone. The proposed system takes advantage of the smartphone audio I/O and processing capabilities to perform acoustic ranging in the audio band using non-invasive audio signals and it has been developed having in mind applications that require high accuracy, such as augmented reality, virtual reality, gaming and audio guides. The system works in a distributed operation mode, i.e. each smartphone is able to obtain its own position using only acoustic signals. To support the positioning system, a Wireless Sensor Network (WSN) of synchronized acoustic beacons is used. To keep the infrastructure in sync we have developed an Automatic Time Synchronization and Syntonization (ATSS) protocol with a standard deviation of the sync offset error below 1.25 μs. Using an improved Time Difference of Arrival (TDoA) estimation approach (which takes advantage of the beacon signals' periodicity) and by performing Non-Line-of-Sight (NLoS) mitigation, we were able to obtain very stable and accurate position estimates with an absolute mean error of less than 10 cm in 95% of the cases and a mean standard deviation of 2.2 cm for a position refresh period of 350 ms.

Contents

List of Figures

List of Tables

Acronyms

1D	One-dimensional space
2D	Two-dimensional space
3D	Three-dimensional space
A/D	Analog-to-Digital Converter
AoA	Angle of Arrival
API	Application Programming Interface
AWGN	Additive White Gaussian Noise
BPSK	Binary Phase Shifting Key
CDMA	Code Division Multiple Access
COTS	Commercial off-the-shelf
CR	Pulse Compression Ratio
D/A	Digital-to-Analog Converter
DBPSK	Differential Binary Phase Shift Keying
DOP	Dilution Of Precision
FDMA	Frequency Division Multiple Access
FFT	Fast Fourier Transform
FIR	Finite Impulse Response
FP	Fingerprinting
FPGA	Field Programmable Gate Array

GDOP	Geometric Dilution of Precision
GNSS	Global Navigation Satellite System
GPS	Global Positioning System
GPU	Graphics Processing Unit
HDOP	Horizontal Dilution of Precision
IMU	Inertial Measurement Unit
IPS	Indoor Positioning System
LAN	Local Area Network
LLS	Linear Least Squares
LoS	Line of Sight
MAC	Medium Access Control
ML	Maximum Likelihood
MT-TDoA	Multiple-Transmitter Time Difference of Arrival
NLLS	Non-Linear Least Squares
NLoS	Non-Line of Sight
PAR	Peak-to-Average Ratio
PC	Personal Computer
PCB	Printed Circuit Board
PDA	Personal Digital Assistant
PDF	Probability Density Function
PDOP	Position Dilution of Precision
PDR	Pedestrian Dead Reckoning
PSL	Peak-Sidelobe Level
RF	Radiofrequency
RFID	Radiofrequency Identification
RIR	Room Impulse Response
ROC	Receiver Operating Characteristic
RSS	Received Signal Strength
RToF	Round-trip Time of Flight
SNR	Signal-to-Noise Ratio

SoC	System-on-Chip
ST-TDoA	Single-Transmitter Time Difference of Arrival
TBP	Time Bandwidth Product
TDMA	Time Division Multiple Access
TDoA	Time Difference of Arrival
TDOP	Time Dilution of Precision
ToA	Time of Arrival
ToF	Time of Flight
UDP	User Datagram Protocol
US	Ultrasound
USART	Universal Synchronous Asynchronous Receiver/Transmitter
USB	Universal Serial Bus
UWB	Ultra Wide Band
VDOP	Vertical Dilution of Precision
WiFi	IEEE 802.11 Wireless Local Area Network
WoI	Wake-on-Interrupt
WSN	Wireless Sensor Network

CHAPTER 1

Introduction

Accurate head positioning is important for indoor augmented reality applications. Most of the current systems use optical tracking methods, i.e. marker-based systems [1][2] or image-based systems [3][4], based on image processing techniques [5][6][7] to obtain the head position in space and its relative orientation. These tracking methods have in general high update rates and relatively short latencies.

However, they are negatively impacted by the well known line-of-sight (LoS) problem. Any obstacle between the camera and the object that is being tracked, seriously degrades the tracking performance [4] and compromises the use of the system in multi-user scenarios. Other factors such as ambient light variation and additional infrared radiation also affect the system performance and are normally mitigated with a carefully designed and lighted environment, to reduce the uncertainty.

In this work we have developed a high accuracy multi-user head tracking system fully compatible with conventional smartphones. The system uses the built-in hardware and processing capabilities of a conventional smartphone to obtain the user head position in a specific room.

1.1 Purpose of Research

This work addresses the indoor multi-user head tracking problem and contributes to a robust, accurate and low-cost solution to perform indoor localization of multiple users who are only carrying a conventional smartphone.

1.1.1 Work Context and Motivation

This work has been developed in the context of the projects: Pervasive Tourism and AcousticAVE - Auralisation Models and Applications in Virtual Reality Environments. The main goal was to

develop a multi-user and high accuracy head tracking system fully compatible with conventional smartphones to be later used as the base of an augmented reality audio guiding system for museums. Conceptually, the overall system enables users to experience binaural audio according to their position and head orientation in relation to several virtual audio sources placed at points of interest in a specific room, thus creating the illusion that a real object acts as a sound source.

Figure 1.1 illustrates the concept of an immersive audio guiding system for museums, that is capable of generating virtual binaural sound sources in real-time. The goal is to create the illusion that the exhibits (e.g. paintings, sculptures or other works of art) emit sound. This would allow visitors walking through a particular room, carrying only headphones and simple, small and comfortable tracking equipment to receive appropriate audio information according to their position and head orientation. The tracking equipment can be as simple as a common smartphone, with its embedded position tracking hardware, connected to an IMU attached to the headphone set for head-tracking.

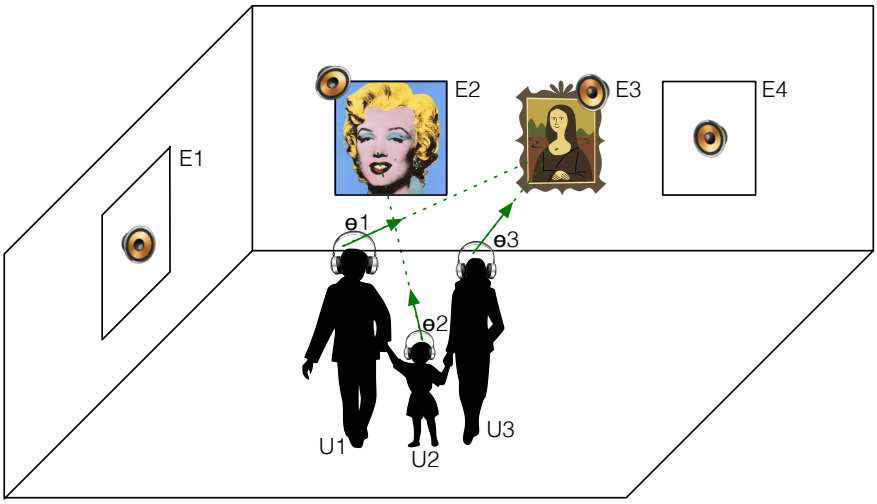


Figure 1.1: Example of the audio augmented reality concept used on a guiding system for museums. Each user carries a tracking device that is capable of generating binaural audio according to head position and orientation in relation to the virtual audio sources assigned to each of the exhibits.

In the example presented in Figure 1.1, users U_1 and U_3 are both looking at exhibit E_3 from distinct positions and the user U_2 is looking at exhibit E_2. All users share the same physical world but with distinct relative head positions and orientations. Conceptually, the goal is to deliver specific binaural audio content to each user in an augmented reality approach by placing a virtual sound source at the exhibit position. For realistic binaural audio generation, the conceptual application introduced before needs high accuracy head positioning (in the centimetre-level order) and accurate heading (θ_j) estimation to generate a real-time and realistic immersive audio experience.

The user head orientation can be obtained using an external Inertial Measurement Unit (IMU). However, we have restricted ourselves to the centimeter-level head positioning problem and we did not considered heading information in this work.

1.1.2 Objectives

The ability to locate objects and people indoors remains a substantial challenge [8] and effective centimeter-level smartphone-based indoor positioning is still an open problem. The research here presented focused in centimeter-level accuracy positioning applications. The purpose of this work was to find a way of estimating the position of a user carrying a conventional smartphone without special hardware.

To achieve the centimeter-level positioning accuracy, Time-of-Flight (ToF) ranging is the most convenient method [9]. As we want to use smartphones as mobile nodes, several design constraints emerge. Commercially available smartphones allow a maximum audio sampling rate of 44.1 kHz therefore limiting the useful band to the Nyquist frequency, i.e. 22.05 kHz. Considering the available hardware inside a conventional smartphone, the only possibility of performing ToF-based ranging is to use the built-in microphone/speaker pair to perform acoustic ranging.

Given all these constraints we conducted our research based on the following questions:

- Is it possible to use the audio band to perform effective non-invasive acoustic ranging? How far is it possible to reduce its perceptual impact on humans?
- Is it possible to effectively estimate the position of a conventional smartphone (without special hardware) based on non-invasive acoustic ranging?
- Since lower frequency content will be used, multipath and interference will increase considerably due to the lower attenuation that signals experience. Is it possible to effectively mitigate these problems to increase the performance of the positioning system?

1.1.3 Contributions

The main contributions of this work can be summarized as follows:

- A methodology for acoustic pulse design with reduced perceptual impact on humans. Two approaches are proposed, the first uses a time domain weighting approach and the second uses a time domain perceptual filter to pre-process the pulses. The perceptual filter design was based on the threshold of hearing data estimated by Suzuki and Takeshima in 2004 [10].
- A method for acoustic beacon synchronization over a WSN infrastructure. The proposed synchronization method is a simplified implementation of the IEEE 1588 standard.
- A robust method for improved smartphone MT-TDoA estimation with NLoS mitigation.
- A simplified model of the indoor acoustic environment for simulation purposes and a simplified acoustic link budget analysis methodology.
- A fully functional prototype using a WSN infrastructure of synchronized acoustic beacons and a positioning App running on a smartphone.

The research behind this thesis work led to the following publications:

Journals indexed by ISI:

- **S. I. Lopes**, J. M. N. Vieira, J. Reis, D. F. Albuquerque and N. B. Carvalho, "Accurate Smartphone Indoor Positioning using a WSN Infrastructure and Non-Invasive Audio for TDoA Estimation". Pervasive and Mobile Computing Journal (2014), Elsevier, ISSN 1574-1192, `http://dx.doi.org/10.1016/j.pmcj.2014.09.003`, **IF: 1.667**.
- D. Albuquerque, J. M. N. Vieira, **S. I. Lopes**, C. A. C. Bastos and P. J. S. G. Ferreira, "Indoor Acoustic Simulator for Ultrasonic Broadband Signals with Doppler Effect", Journal of Applied Acoustics, Elsevier, ISSN 0003-682X, accepted in December 2014, **IF: 1.068**.

International Conferences:

- **S. I. Lopes**, J. M. N. Vieira, D. Albuquerque, "Analysis of the Perceptual Impact of High Frequency Audio Pulses in Smartphone-based Positioning Systems". 2015 IEEE International Conference on Industrial Technology, Seville, Spain, 17-19th March 2015, accepted in December 2014.
- T. Aguilera, F. J. Álvarez, A. Sánchez, D. Albuquerque, J. M. N. Vieira and **S. I. Lopes**, "Characterization of the Near-Far Problem in a CDMA-Based Acoustic Localization System". 2015 IEEE International Conference on Industrial Technology, Seville, Spain, 17-19th March 2015, accepted in December 2014.
- G. Campos, P. Dias, J. M. N. Vieira, J. Santos, C. Mendonça, J. P. Lamas, N. Silva, **S. I. Lopes**, "Acousticave: Auralisation Models and Applications in Virtual Reality Environments". 45^{o} Congreso Español de Acústica, 8^{o} Congreso Ibérico de Acústica. European Symposium on Smart Cities and Environmental Acoustics, 29-31 October, Murcia, Spain.
- **S. I. Lopes**, J. M. N. Vieira, G. Campos and P. Dias, "Sistema de Realidade Aumentada Áudio 3D para Dispositivos iOS". In Proceedings of the 12^{th} Audio Engineering Society Brazil Conference (AES Brazil 2014); São Paulo, Brazil, 13-15 May, 2014, ISSN 2177-529X.
- **S. I. Lopes**, J. M. N. Vieira, D. F. Albuquerque and N. B. Carvalho. "Accurate Smartphone Indoor Positioning Using Non-Invasive Audio". In Proceedings of the 2013 International Conference on Indoor Positioning and Indoor Navigation (IPIN'2013). Montbeliard, France, 28-31 October 2013.
- D. F. Albuquerque, J. M. N. Vieira, **S. I. Lopes**, C. A. C. Bastos and P. J. S. G. Ferreira. "OFDM Pulse Design with Extremely Low PAPR for Ultrasonic Location and Positioning Systems". In Proceedings of the 2013 International Conference on Indoor Positioning and Indoor Navigation (IPIN'2013). Montbeliard, France, 28-31 October 2013.

- **S. I. Lopes**, A. Oliveira, J. M. N. Vieira, G. Campos, P. Dias and R. Costa, "Real-Time Audio Augmented Reality System for Pervasive Applications". In Proceedings of the 11^{th} Audio Engineering Society Brazil Conference (AES Brazil 2013); R.R.A.Faria (ed.); São Paulo, Brazil, 7-9 May, 2013, ISSN 2177-529X.

- **S. I. Lopes**, J. M. N. Vieira and D. Albuquerque, "High Accuracy 3D Indoor Positioning Using Broadband Ultrasonic Signals". Trust, Security and Privacy in Computing and Communications (TrustCom), 2012 IEEE 11th International Conference on, pp.2008-2014, 25-27 June, 2012, `http://doi.ieeecomputersociety.org/10.1109/TrustCom.2012.172`

- D. F. Albuquerque, J. M. N. Vieira, **S. I. Lopes**, C.A.C. Bastos and P. J. S. G. Ferreira, "Indoor Ultrasonic Transfer Function for Moving Objects". In Proceedings of the 2011 International Conference on Indoor Positioning and Indoor Navigation (IPIN'2011); Moreira, Adriano J. C.; Meneses, Filipe M. L. (eds.), Guimarães, Portugal, 21-23 Setembro 2011, ISBN: 978-972-8692-63-6.

- D. F. Albuquerque, J. M. N. Vieira, **S. I. Lopes**, C. A. C. Bastos and P. J. S. G. Ferreira, "Indoor Ultrasonic Simulator for Moving Objects Using Delay Filters". In Proceedings of the 2011 International Conference on Indoor Positioning and Indoor Navigation (IPIN'2011); Moreira, Adriano J. C.; Meneses, Filipe M. L. (eds.), Guimarães, Portugal, 21-23 Setembro 2011, ISBN: 978-972-8692-63-6.

Poster Sessions:

- **S. I. Lopes**, J. M. N. Vieira, N. B. Carvalho. "Accurate Smartphone Indoor Positioning using a WSN Infrastructure and Non-Invasive Audio for TDoA Estimation". Research Day 2014, Universidade de Aveiro, 3 de Junho, 2014.

- **S. I. Lopes**, J.M.N. Vieira, "Audio Augmented Reality System for Museums", Bolsa de inovação em Turismo da 2^{a} Conferência Internacional INVTUR 2012. Universidade de Aveiro, 16 a 19 de Maio, 2012.

1.2 Definition of Terms

Given the quantity and complexity of the information related to indoor user head tracking and positioning topics, we chose to add at this point the definition of a set of terms to improve the readability of this document.

We followed the definitions introduced by Rainer Mautz in [8] (Section 3.1), to disambiguate some important terms that are often used in the literature, such as positioning, localization, navigation and tracking.

Additionally, we define the following terms that are recurrently used throughout this document:

High-accuracy - we use this term to classify systems with centimeter-level accuracy.

Node - any device that is part of a positioning network. Nodes can be infrastructure elements such as beacon nodes, or target nodes such as: smartphones, tablets or laptop computers.

Beacon node - also known in the literature as anchor nodes or landmarks. In this document we refer to a beacon as a transmitter with a known and static position.

Target node - also known in the literature as mobile nodes or mobile devices. In this document we refer to a target node as a receiver with unknown position.

Self-positioning - also known in literature as GPS-like or distributed positioning systems. In this positioning topology the target node estimates its own position based on the measured signals that it receives from beacon nodes.

Non-invasive - we use this term to define signals with audio and ultrasonic spectral content, but with pre-processing included to make the audio band non-invasive, i.e. with reduced perceptual impact on humans. An acoustic ranging classification related to the frequency bands of the acoustic signals, is defined in Figure 1.2. Acoustic ranging can be performed using audio or ultrasound signals. In this work we extend the non-invasive area to include the upper part of the audio band, i.e. above 15 kHz.

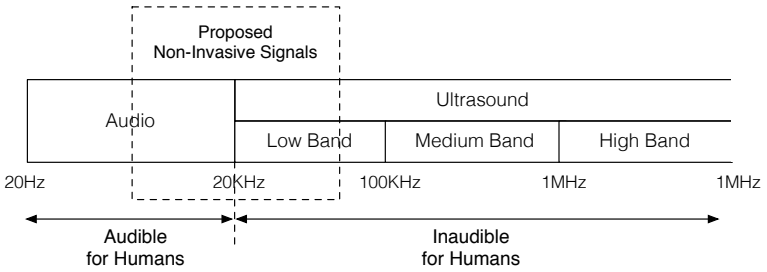


Figure 1.2: Non-Invasive signals classification.

1.3 Design Challenges

Follows a description of the main challenges that should be considered in the application design process:

1. Centimeter-level Accuracy: To satisfy this criterion a ToF range-based positioning system should be used. In addition, an infrastructure of synchronized beacons should be considered to circumvent the lack of accuracy of mutual positioning systems [9].

2. Smartphone Compatibility: The proposed system must be compatible with conventional smartphones. This requirement restricts the selection of the sampling frequency of the acoustic signal used in ranging, due to smartphone hardware constraints. Commercially available smartphones allow a maximum sampling rate of 44.1 kHz therefore limiting the useful acoustic band to the Nyquist frequency, i.e. 22.05 kHz.
3. Non-Invasive Pulse Design: The signals proposed should be carefully designed to make them non-invasive, i.e. with reduced perceptual impact on humans. An important challenge is how to reduce the transmitted power while maintaining the detection sensitivity.
4. Low-Cost Infrastructure: A WSN must be used with a low beacon node density. System infrastructure must be based on low-cost and comercial-off-the-shelf components. An important challenge is related to the synchronization strategy that will be used to keep the network elements in sync to perform ToF range measurements with increased accuracy. An automatic time synchronization protocol to keep the WSN nodes in sync should be considered.
5. Multi-User Approach: One-way passive ranging (i.e. GPS-like) should be used in a self-positioning architecture to maximize the multi-user capacity. Additionally, the WSN must use a star topology per room and rooms must be interconnected through a main router to the internet.

1.4 Thesis Structure

This thesis is organized in seven chapters. Chapter 2 provides an overview of the technological diversity of positioning systems, with a focus on methods and techniques commonly used in indoor positioning and tracking applications. Additionally, a state-of-the-art in centimeter-level Indoor Positioning Systems (IPS) is presented.

Chapter 3 focuses on the non-invasive acoustic ranging topic. It starts with an overview of the acoustics theory that is used throughout this thesis to model the indoor acoustic environment. A study on the perceptual impact of sounds on humans is taken, starting with the human hearing system characterization, i.e. frequency response and loudness sensation, to propose a series of non-invasive pulses that will be used for ranging. In the pulse design stage, two main approaches are then presented: Time Domain Weighting and Perceptual Filtering. Lastly, we describe the experimental procedure used for the perceptual evaluation of the proposed pulses.

In Chapter 4 we propose the design of an infrastructure of synchronized acoustic beacons to enable ToF measurements, and thus obtain high accuracy position estimates. First, coordination methods are discussed in relation to the medium access; then the proposed synchronization strategy is addressed and lastly the acoustic beacon prototype design is described and its experimental validation is performed.

Chapter 5 describes the overall head positioning system architecture. Firstly is introduced the architecture of the proposed system followed by a detailed description of the measurement and

positioning approaches used. In the end of the chapter an experimental validation is conducted using an iPhone 4S to evaluate the positioning system in real applications. Two experiments were done to evaluate the system performance. The first experiment was managed to obtain a quantitative evaluation of the overall system by measuring the position estimated in a grid of points in a regular room and a second experiment was conducted to obtain a qualitative evaluation of the positioning system when a person equipped with a mobile device is in a moving trajectory.

Finally, in Chapter 6 the results obtained in Chapters 3, 4 and 5 are evaluated and discussed and in Chapter 7 concluding remarks are given together with recommendations and ideas for future research.

CHAPTER 2

Indoor Positioning and Tracking

GNSS (Global Navigation Satellite Systems) is the most widely used method for outdoor localization and it provides global coordinates with an accuracy within 10 meters. GNSS signals can also be used indoors, but special hardware and powerful digital signal processing techniques are required to extract the Time-of-Flight (ToF) information from the noisy signals before a position estimate can be computed [11] [12]. However, due to their low accuracy, GNSS are not suited for indoor centimeter-level applications. In this chapter, we address the technological diversity in positioning systems with a focus on the technologies and current systems that are most related to the scope of this work. At the end of this chapter we introduce the positioning process and propose a simplified taxonomy to organize the most important methods and techniques used for indoor centimeter-level positioning.

2.1 Technological Diversity in Indoor Positioning Systems

In [8] Rainer Mautz grouped the technologies commonly used for positioning. Figure 2.1 depicts the most common technologies for user tracking in terms of their accuracy and coverage area. By restricting the accuracy and coverage in Figure 2.1 to meet the indoor scenario (i.e. coverage of 50 m), and focusing on the head tracking application case (i.e. accuracy below 1 m) only ten technologies distributed by five categories became eligible to implement an effective Indoor Positioning System (IPS) for user tracking applications.

In Table 2.1 these technologies are grouped by category and we will mainly focus on centimeter-level accuracy, due to the constraints imposed by our application case. From the Table 2.1, an important observation is that effective IPS can be divided into five main categories: 1) Optical-based; 2) Mechanical-based; 3) Magnetic-based; 4) Acoustic-based and 5) RF-based systems.

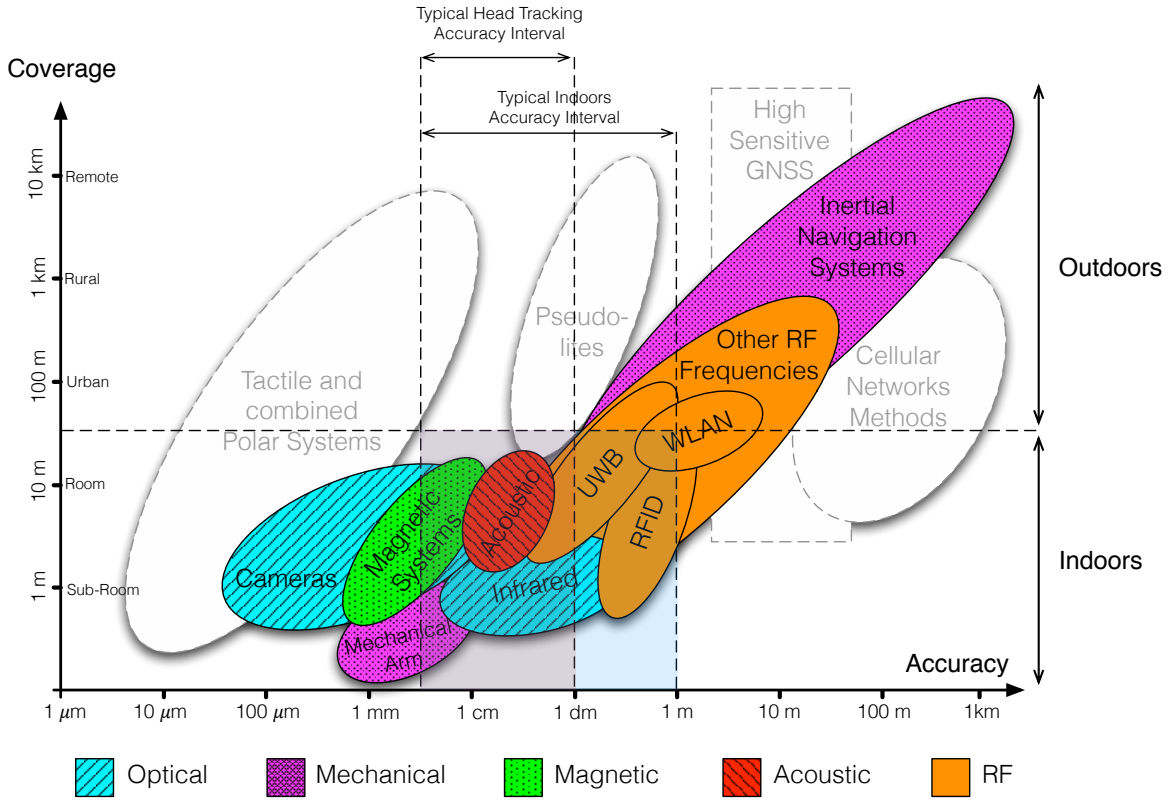


Figure 2.1: Overview of positioning technologies. Adapted from [8].

Category	Technology	Accuracy (m)	Coverage (m)
Optical	Cameras	0.1 mm – dm	1 – 10
	Infrared	mm – cm	1 – 10
Mechanical	Mechanical Arm	0.1 mm – 1 mm	1 – 2
	Inertial Navigation	1 %	10 – 100
Magnetic	Magnetic Systems	mm – cm	1 – 20
Acoustic	Sound	cm	2 – 10
RF	UWB	cm – m	1 – 50
	RFID	dm – m	1 – 50
	WLAN	m	20 – 50

Table 2.1: Overview of indoor system accuracy grouped by technology. Adapted from [8].

In the following subsections these five categories will be introduced and discussed in terms of its principles of operation, related works and commercial availability. In the end, a discussion will take place to point major advantages and drawbacks of each of these categories with focus in the following criteria: ranging method, reported accuracy, simultaneous users, bandwidth, update rate, cost and smartphone compatibility.

2.1.1 Optical-based Systems

This type of systems integrates cameras and infrared technologies and can be classified in two main types: 1) marker-based and 2) image-based systems. Marker-based systems [1] [2] use active or passive markers (e.g. using reflective materials) attached to the object that is being tracked, and one or multiple cameras to detect the marker positions when infrared light is being used. In these conditions the image processing algorithm is reduced to a simple blob detection in a grayscale image, see the Vicon Tracking system example in Figure 2.2. These type of systems are normally used for human body tracking, in the movie industry, and are very expensive.

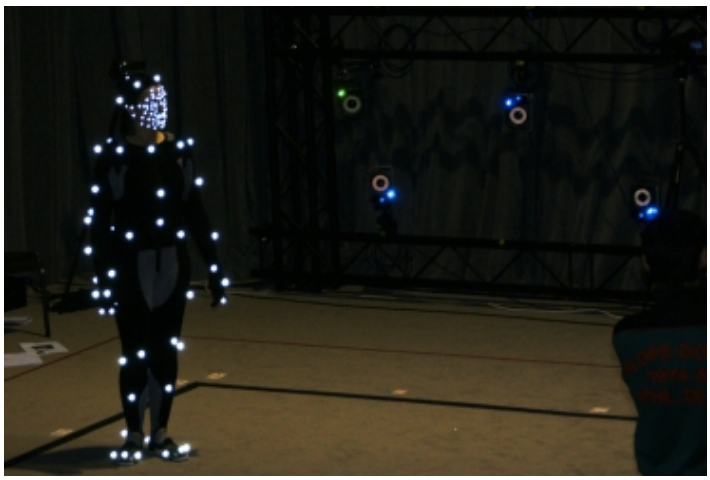


Figure 2.2: Marker-based Vicon tracking system. Image from [13].

Image based systems [3][4] tend to be more complex due to the increase of image processing needs. This type of systems use image processing techniques to obtain features of the objects to be tracked. These features are then used for face, eye, or head recognition, which enables the use of tracking algorithms, e.g. Kalman and particle filters. This type of systems can be based on a single camera [14], or dual cameras [15][5] to obtain depth information by combining two 2D images taken at slightly different points, cf. Figure 2.3.

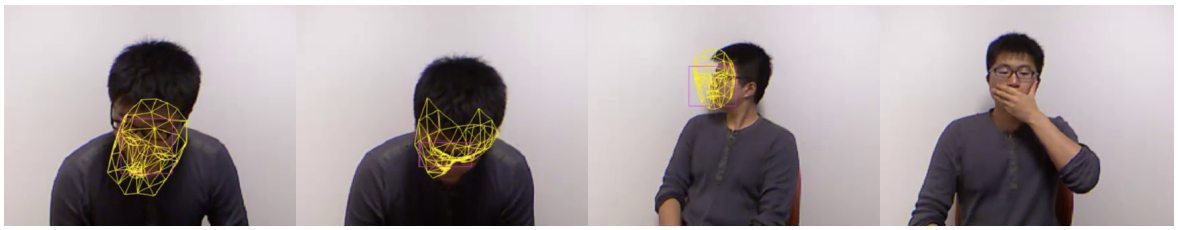


Figure 2.3: Image-based tracking using Microsoft Kinect. Image from [16].

Major advantages and drawbacks of the optical-based systems are:

Advantages:

- High accuracy and high update rate;
- High availability;
- Can work over a large area;
- No magnetic interference problems.

Drawbacks:

- Heavy LoS dependence;
- Limited by intensity and coherence of light sources;
- Complex system infrastructure.

2.1.2 Mechanical-based Systems

The Mechanical-based systems group is divided in two main types: 1) Mechanical Arm Systems and 2) Inertial-based Systems.

Mechanical Arm systems use a direct mechanical connection (a light-weight arm equipped with multiple encoders and a control system) between a reference point and the head that is being tracked, cf. Figure 2.4. This type of system measures the changes in position and orientation with respect to a reference point.

Figure 2.4: Mechanical arm head tracking system from Fakespace. Image from [17].

Major advantages and drawbacks of mechanical arm systems are:

Advantages:

- High accuracy, high update rate and low latency;
- Low cost and robustness.

Drawbacks:

- Single-user tracking;
- Intrusiveness due to the existence of a physical connection thus resulting in strong limitations imposed on user movements [18].

Inertial-based methods use Inertial Measurement Units (IMU) to obtain estimates of position, velocity and orientation of a user. An IMU (Figure 2.5) is composed by three orthogonally arranged accelerometers (motion sensors), three gyroscopes (angular rate sensors) and/or a magnetometer (3 perpendicular sensors for measuring the strength and/or direction of a magnetic field) [8]. By using sensor fusion techniques (i.e. Kalman filters or Particle filters) the IMU data can be used to obtain the direction/orientation (in the literature it is also known as yaw or heading) of the moving target.

Figure 2.5: Inertial tracking systems from Intersense. InertiaCube BT (bluetooth version) InertiaCube 4 (USB version). Images from [19].

Major advantages and drawbacks of inertial-based systems are:

Advantages:

- High update rate and low latency;
- Multiple user tracking (each user carries his/her own device);
- Unlimited range;
- No LoS problem;
- Small size and low cost.

Drawbacks:

- Accuracy is velocity dependent;
- Significant error drift over time, i.e. error integration causes an accuracy degradation over time [20].

If the initial position and orientation are known and high precision IMU sensors are used, it is possible to perform double integration to update the velocity and position of the moving target. This method has significant error drift over time due to the limited precision of the IMU sensors, thus limiting its use in real world applications. All the data directly extracted from the IMU and obtained with the sensor fusion techniques can be used for tracking with the Pedestrian Dead Reckoning (PDR) method. Dead reckoning estimates a position based upon previously determined positions and known or estimated speeds over the elapsed time. In the literature, PDR is used as an indication that accelerometers have been attached to the body of a person [8].

2.1.3 Magnetic-based Systems

Magnetic tracking systems, e.g. [21, 22], typically use measurements of the strength of a 3D magnetic field generated by three small wire coils placed perpendicular to each other. These three coils are then attached to the object to track. By sequentially changing the 3D magnetic field transmitted by the coils, it is possible to obtain the position and orientation of the object.

Figure 2.6: Magnetic tracking system from Polhemus. Image from [23].

Major advantages and drawbacks of magnetic-based systems are:

Advantages:

- Multiple user tracking;
- No LoS problem.

Drawbacks:

- Accuracy is highly dependent on external magnetic interferences, as well as metal objects;
- Coverage area tends to be small.

2.1.4 Acoustic-based Systems

Acoustic-based systems use sound or ultrasound [24, 25] to measure the position and orientation of the target object. There are two ways of doing it, using the time-of-flight or the phase-coherence method.

Time-of-flight (ToF) (also referred in the literature as ToA, cf. Section 2.3.1.2) is an absolute tracking method and works by measuring the time that it takes for sound to travel from the transmitters (at fixed positions) to the receiver (object to track). Using a multiple access medium technique it is possible to obtain independent distances from the receiver to each of the transmitters and thus compute a position estimate using trilateration, cf. Section 2.3.2.2. Orientation can also be obtained if a sensor array is used at the receiver to estimate the angle-of-arrival.

Major advantages and drawbacks of ToF acoustic systems are:

Advantages:

- High accuracy and low latency;
- Multiple user tracking;
- Electromagnetic interference immunity;
- LoS can be easily mitigated;
- Low cost.

Drawbacks:

- Low update rate;
- Considerable channel multipath and fading variation over time.

Phase coherence is a relative tracking method, cf. [26], that measures the difference in phase between sinusoidal sound waves generated by a transmitter on the receiver (object to track) and those emitted by a transmitter at some reference point. As long as the distance travelled by the receiver is less than one wavelength between updates, the system can update the position of the receiver. By using multiple transmitters orientation can also be determined.

Major advantages and drawbacks of phase coherence acoustic systems are:

Advantages:

- Low latency;
- Electromagnetic interference immunity;
- Multiple user tracking;
- Low cost.

Drawbacks:

- Time-varying accuracy, as the error tends to be accumulated over time due to the relative tracking nature of the method;
- Considerable channel multipath and fading variation over time.

2.1.5 RF-based Systems

RF-based systems, such as WiFi/Fingerprinting (FP), RFID and Ultra-Wideband (UWB) are well known and widely used methods.

WiFi/Fingerprinting, are commonly used in low accuracy applications and employ an off-line calibration process to build a map of "fingerprints" that are empirically obtained by taking exhaustive RSS measurements inside a building. In operation a position estimate is obtained by matching online measurements with the pre-computed "fingerprints" available in a database, cf. Sections 2.3.1.1 and 2.3.2.1.

RFID-based systems commonly integrate RFID tags in the pavement or in the walls. The principle for positioning is based on proximity, i.e. target nodes need a built in transceiver to scan the presence of a unique RFID tag that is geo-referenced in the building map where the user is moving [8].

UWB positioning systems use narrow pulses with very short duration (subnanosecond) resulting in widely spread radio signals in the frequency domain [27] and in Time of Arrival (ToA) measurements with increased accuracy, when compared with other RF methods [28] [8]. However, UWB systems need high accuracy timing resulting in complex and costly hardware. In range-based positioning systems, time measurements are used to estimate the ToA or the Time-Difference-of-Arrival (TDoA) to compute a position estimate [29], with the resulting accuracy being directly proportional to the signal bandwidth. UWB systems normally present heavy timing constraints only reachable using specific synchronization methods to keep all the intervenient nodes in sync, i.e. under the same universal clock, thus resulting in an expensive system infrastructure.

2.2 State-Of-The-Art in Centimeter-Level Positioning Systems

In this section we restrict ourselves to centimeter-level positioning systems, which comprise systems from both acoustic and RF categories. The focus on these categories is justified by the design decisions we took in the technology selection process mainly due to the requirements imposed by the scope of this work, from which we highlight: (1) Centimeter-level Accuracy, (2) Low-Cost Infrastructure, (3) System Scalability and (4) Smartphone Compatibility.

We discuss six systems from the acoustic category: Guoguo [30]; LOSNUS [31]; 3D Locus [32]; Parrot [33]; Cricket [34, 35] and Beep [36]. We also discuss six systems from the RF category: NOKIA HAIP [37]; Ubisense UWB [38]; Pietrzyk et al. UWB Experimental Plataform [42]; Segura et al. UWB System [43]; RADAR [39] and COMPASS [40].

2.2.1 Guoguo

The Guoguo indoor positioning system [30] is an infrastructure-based system that uses smartphones as mobile stations for accurate indoor positioning. The system uses a network of beacons (anchors)

with a coordination protocol to transmit modulated pulses using audible-band acoustic signals (of less than 20 kHz), a real-time processing app running in a smartphone, and a backend server for indoor contexts and location-based services [30], cf. Figure 2.7.

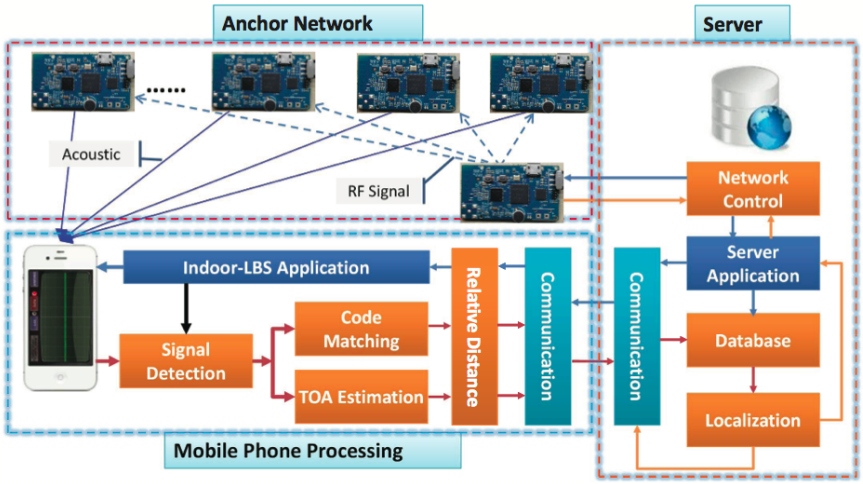


Figure 2.7: Conceptual Architecture of the Guoguo System. Image from [30].

The beacons transmit modulated acoustic pulses synchronized at the instants determined by the Network Control server side block. The synchronization process is performed using radio tokens through a Zigbee RF module. These radio tokens are transmitted independently for each beacon in the network. When a beacon detects its specific token, it begins transmitting the acoustic pulse.

The smartphone detects and decodes the data embedded in the acoustic pulses and computes the relative ToA for each beacon, i.e. the MT-TDoA. In [30] the authors state that the Guoguo system is able to locate a smartphone at the centimeter-level with an average localization accuracy below 25 cm in typical office or classroom environments.

The authors claim that the system is "unnoticeable to humans" and present a preliminary study based only on acoustic background noise measurements with the system working. They did not evaluate the signals with subjects, being their classification of "unnoticeable to humans" signals unfounded.

2.2.2 LOSNUS

The LOSNUS (LOcalization of Sensor Nodes by Ultra-Sound) [31] system offers 3D high accuracy in the centimeter order, with a refresh rate of up to 10 Hz and can be used for both tracking and locating static devices. The system uses a fixed infrastructure of six US transmitters (Polaroid 600) with known positions, mounted in a room wall and connected through coaxial cables to a centralized demultiplexer to activate the transmitters. The mobile devices are nodes that form a WSN, equipped with a MEMS US receiver (SPM0204UD5 from Knowles Acoustics) and a microcontroller equipped with a RF link for communications [31].

The system uses an optimized sequence of coded US signal frames transmitted by different transmitters [31] to improve the position refresh rate and reduce signal interference between the transmitted signals. The transmitters are sequentially activated with a linear frequency modulated chirp signal with bandwidth 35 kHz to 65 kHz for ToF measurement followed by the transmission of an unique orthogonal pseudo-random code [31].

2.2.3 3D-LOCUS

The 3D-LOCUS [32] is a general centralized positioning system that uses sound (5 to 25 kHz) for ranging (using ToF) and RF (Bluetooth for data and the 433 MHz band for synchronization) or wires (BusCAN for data and LVDS for synchronization). The 3D-LOCUS can operate using two different access modes: time multiplexing (TDMA) or code multiplexing (CDMA). In both cases a direct sequence spread spectrum is used (Golay codes with length of 32, 64 or 128) modulated with BPSK. In the normal configuration the system has a network of sensors mounted on the ceiling, connected by wires to a central node, and some moving tags, which are interconnected by RF. The system can work in three different configurations: Centralized (the moving tags send an ultrasonic pulse and the central node estimates their position), Privacy Oriented (the ceiling tags send an ultrasonic pulse and each moving tag estimates its position) or Bidirectional (both type of tags send an ultrasonic pulse). In addition to position estimation, the system is also able to estimate the speed of sound.

The system has an accuracy of 4.1 to 5.2 mm in 90% of the cases for TDMA access mode and 8.6 to 11 mm for CDMA mode; with a wind flow of 2 m/s the system accuracy changes to 4.9 to 11.5 mm in 90% of the cases for TDMA access mode and 7.5 to 13.7 mm for CDMA mode. Therefore, this system can operate not only in indoor environments but also in outdoor environments. However, the system only operates in the range from 1.02 to 4.94 m which limits the system applicability. Moreover, this system requires a calibration process to achieve the sub-centimeter accuracy.

2.2.4 Parrot

The Parrot system [33] is based on a network of nodes (Parrots), cf. Figure 2.8. Each node uses four ultrasonic transmitters and receivers (40 kHz) plus an auxiliary radio channel for synchronization purposes and data communication. The system randomly sorts the nodes in a queue and the first node in the queue sends both ultrasonic and RF pulses simultaneously. The other nodes receive these signals and compute the distance to that node using ST-TDoA measurements. Then, this node goes to the end of the queue and the first node in the queue behaves the same way of the last one, and so on.

The RF link between the nodes is used also to communicate the measured distances from one node to the others. This schema thus eliminates the necessity of a central system to compute the node positions. Each node computes its position locally from the information given by the other

nodes. This positioning system exhibits an accuracy of 2 cm with a standard deviation of 4 cm and it can reach up to 15 m.

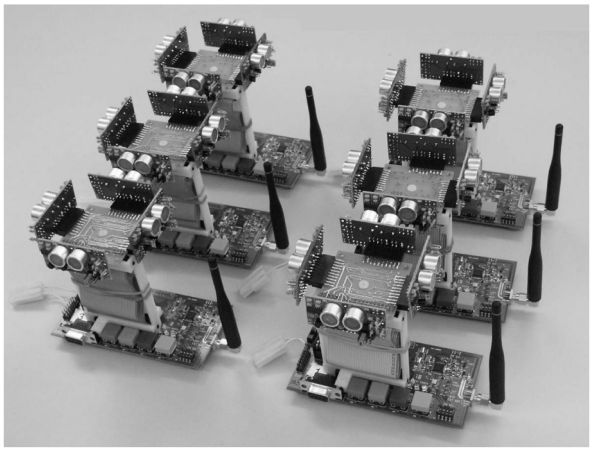


Figure 2.8: Parrot sensor nodes. Image from [33].

2.2.5 Cricket

The Cricket IPS [34, 35] allows a node to localize itself and obtain its own orientation, by using ultrasound and an RF receiver module. The Cricket positioning system uses several ultrasonic and RF transmitters (mounted on the ceiling or on the wall). These transmitters provide the signal reference that allows the node to obtain its own position and orientation. These transmitters randomly send at the same time, a 47 ms RF signal (at 418 MHz) and an 125 μs ultrasonic pulse (at 40 kHz).

The RF signal carries the emitter ID, position, and the room ID. The nodes receive the RF signal and the ultrasonic pulse from each transmitter and compute the distance to them based on ST-TDoA measurements. The Cricket positioning system uses inexpensive hardware (Figure 2.9): each transmitter or receiver costs less than $10 and provides an accuracy of 10 cm. Moreover, the Cricket Positioning System has a high level of scalability, although the system consumes some power (about 15 mW).

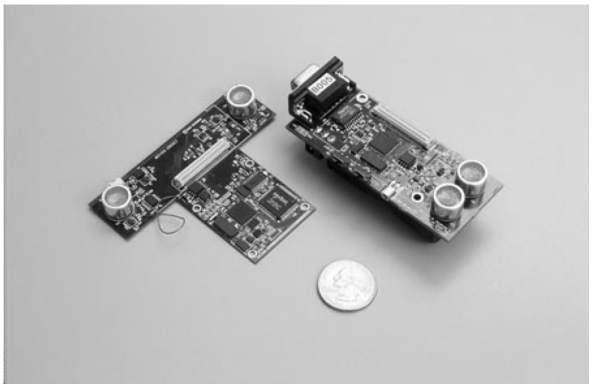


Figure 2.9: Cricket sensor nodes. Image from [41].

2.2.6 Beep

The Beep system is an indoor positioning system that uses audible sound to estimate the position of a mobile device (e.g. cell phones, PDAs, desktops) using an infrastructure of receiver sensors (S_x) with known positions, cf. Figure 2.10. By using audible sound the system can be used easily with existing portable devices without the need of extra hardware. The system uses ToA measurements for ranging and 3D multilateration for position estimation.

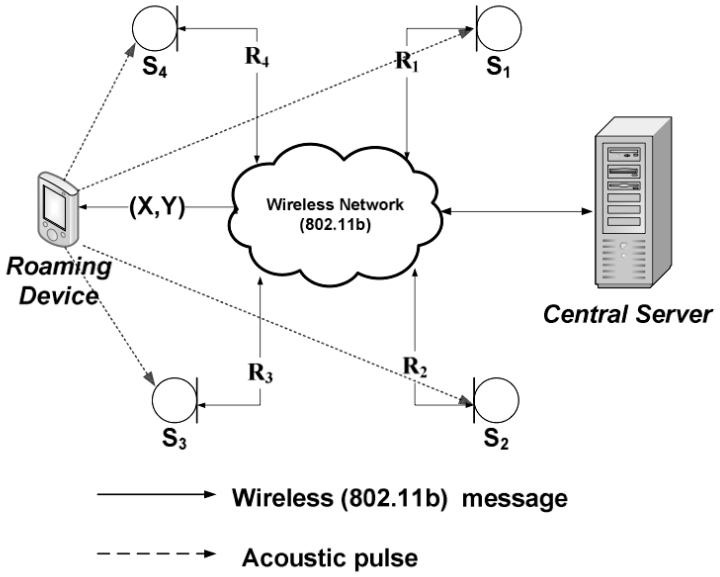


Figure 2.10: Overview of Beep Architecture. Image from [36].

The system infrastructure consists of a set of receiver sensors that are connected to a central server through a WiFi network. Each of these sensors has built-in processing capabilities, a microphone for detecting acoustic signals and a wireless network interface that enables the creation of a local network between all deployed sensors. The synchronization between sensors is achieved through the wireless network connection by exchanging messages which enable the ToA estimation of the transmitted acoustic pulses.

The system works as follows. When a user requests its relative position, the PDA synchronizes with the infrastructure sensors through the wireless network, and transmits a pre-defined acoustic signal. The sensors detect this signal, using specialized digital filters, and compute ToA estimate. This ToA estimate is then translated into distance by multiplying by the speed of sound. The distances obtained by each sensor are then reported to the central server that computes a 3D position estimate using a multilateration approach. Finally the position estimate is reported to the PDA that performed the positioning request. In [36] the authors reported an accuracy of about 61 cm in more than 97% cases.

2.2.7 Nokia HAIP

Nokia HAIP [37] positioning system is built on top of Bluetooth technology and uses an infrastructure of directional receiver antennas (Locators) to obtain AoA measurements. Each Locator is continuously performing AoA measurements from different tags or mobile stations that are continuously transmitting Bluetooth packets at a constant rate, in asynchronous mode. The AoA measurements are then sent to a positioning server via WiFi to estimate the HAIP Tag position based on the AoA obtained measurements. The system uses small HAIP Tags, and ceiling mounted HAIP Locator(s) antennas and a remote HAIP Controller unit to compute the position of the HAIP Tags or mobile devices, cf. Figure 2.11.

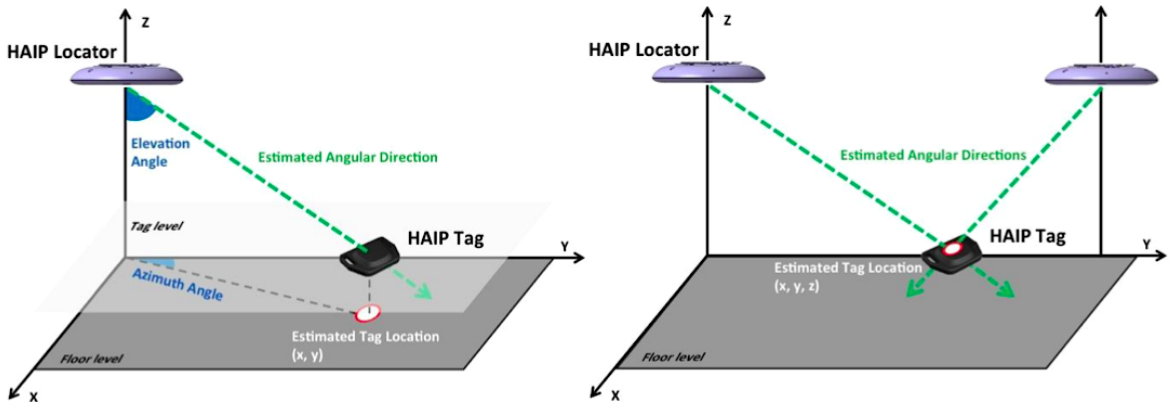


Figure 2.11: Nokia HAIP System used for 2D/3D positioning. Image from [37].

If only a single HAIP Locator is used, azimuth and elevation estimates can be obtained, and if the height of the tag is assumed to be fixed, the HAIP system can estimate the 2D position of the HAIP tag using simple trigonometry, cf. Figure 2.11 at left. If at least two or more HAIP Locators are used, two azimuth and elevation estimates can be obtained which enables the system to compute the 3D position estimate of a HAIP Tag using simple trigonometry, cf. Figure 2.11 at right. The accuracy of the HAIP system depends on the range but with practical ranges between 3 m and 50 m the attained accuracy is in the interval of 30-100 cm.

2.2.8 Ubisense UWB

A commercially available UWB system is presented by Ubisense in [38]. In the datasheet the manufacturer introduces the system as a high accuracy positioning system for industrial applications and claims an accuracy of less than 30 cm even in multipath-rich indoor environments. The Ubisense UWB location system works in a centralized operation mode [38], based in a network of UWB antennas (Figure 2.12) that can be used to measure both the Angle-of-Arrival (AoA) and the Time-Difference-of-Arrival (TDoA) of the pulses transmitted by the mobile station, resulting in accurate

3D tracking information at a rate of 134 measurements per second. Due to the proprietary nature of this system, no information about the internal architecture of the UWB beacons and tags, is available.

Figure 2.12: Ubisense UWB anchor nodes. Image from [38].

2.2.9 Pietrzyk et al. UWB Experimental Plataform

In [42], Pietrzykis et al. propose an experimental UWB-based ranging platform with real-time signal processing. The platform uses ToA for ranging and a non-coherent energy detection receiver architecture. The receiver does not require high sampling rates nor information about the channel. The proposed UWB platform consists of a transmitter, a receiver, and a trigger unit. The transmitter is a Picosecond 3500D pulse generator that is able to produce an ultra short pulse (65 ps of duration) at a repetition rate of 0.5 MHz. The trigger unit that provides the time synchronization between the transmitter and the receiver consists of a Tektronix AWG 7122B Arbitrary Waveform Generator and an Analog Devices AD9516 PLL evaluation board that adjusts the trigger signal in frequency and amplitude as required by the Picosecond generator and the oscilloscope. As a receiver, a LeCroy WE100H sampling oscilloscope with a 30 GHz analog bandwidth is used [42].

All signal processing, i.e. signal squaring, integration, A/D conversion, threshold selection, and distance calculation are performed in real-time at the receiver side by a Visual Basic application operating on the sampling oscilloscope. The authors reported an accuracy interval of 1-2 cm with a sampling rate of 500 MS/s.

2.2.10 Segura et al. UWB System

In [43] the authors propose a UWB system that enables mobile robots to navigate in indoor environments. The system uses several beacons located at fixed positions that transmit synchronized DBPSK modulated pulses with duration of 2 ns. Each mobile robot is equipped with an UWB receiver to obtain MT-TDoA measurements that are then used to derive a position estimate using

a multilateration approach. The proposed prototype uses COTS components and the signal acquisition is performed using an FPGA. The authors report a positioning accuracy below 20 cm in both LoS and NLoS conditions in relation to the fixed beacon nodes in a typical indoor environment [43].

2.2.11 COMPASS

The COMPASS positioning system [40] is a 2D positioning system that uses the installed WiFi infrastructure and a digital compass to obtain the node positions. The COMPASS system is based on the WiFi fingerprints of several known positions and orientations. Therefore, when a node wants to determine its own position it measures the received signal strength of several WiFi access points, and knowing its orientation (given by the digital compass) it is able to compute its position using a probabilistic positioning algorithm, which compares the collected data with the fingerprint data in the same direction.

Practical results showed an accuracy less than 2 m in 70% of the cases (which is enough to track objects with room-level accuracy) in a floor with 15 x 36 m and 9 WiFi access points. The results showed that the introduction of the digital compass makes the position system more accurate and can be useful in dynamic nodes where orientation can help in the prediction of their movements.

Moreover, the system is inexpensive and scalable, uses off-the-shelf components and is compatible with the WiFi installed infrastructures. However, the nodes need to have WiFi hardware, which is a limitation to portability mainly because WiFi hardware significantly increases the weight and power consumption of the nodes.

2.2.12 RADAR

The RADAR positioning system [39] is a 2D positioning and tracking system that uses the wireless LAN in the 2.4 GHz free band (WaveLAN) similar to the WiFi. The system can use the existing wireless LAN infrastructure (the access points) to perform the localization. These access points, with well known positions, broadcast an UDP packet every 4 s. The nodes compute their position by measuring the received signal strength from the access points and by applying the nearest neighbor(s) algorithm. This means that the system selects the position of the node by comparing the received signal strength to a set of a priori signals strengths collected in known locations. These a priori signal strengths can be collected by two different methods: empirical method and radio propagation model method. In the empirical method the set of signal strength is collected beforehand in the real environment. In the radio propagation method, the set of signal strength is computed using a propagation model (wall attenuation factor) of the local.

The RADAR system produces a result with an accuracy of less than 2.94 m in 50% of the cases and less than 4.69 m in 75% of the cases, using the empirical method in 70 places on a floor with 44 x 23 m and 3 access points. Moreover, using the radio propagation model method, the system results showed an accuracy less than 4.3 m in 50% of the cases in the same floor. Therefore, the

accuracy of the system is enough for objects tracking with room-level accuracy.

The system is inexpensive and scalable and it uses COTS components. However, the nodes need to have WiFi capabilities, which is a limitation to portability mainly because the WiFi hardware significantly increases the weight and power consumption of the nodes.

2.2.13 Comparative evaluation of the studied systems

The positioning systems introduced before are grouped by category and compared in Table 2.2. The metrics and features used for comparison were selected having in mind the scope of the proposed work, namely, accuracy and smartphone compatibility.

From Table 2.2 we observe that most of the selected acoustic systems have centimeter-level accuracy and all use time-based methods in the ranging process thus reaching generally the best performance in terms of accuracy among all the presented systems. In relation to the smartphone compatibility feature, only the Guoguo and the Beep systems are eligible.

RF systems based in WiFi/FP showed accuracies below 2 m in 70% of the cases (COMPASS system) and below 4,7 m in 75% of the cases (RADAR system), which is far from acceptable for our application case.

The UWB approaches (UBISENSE, Pietrzyk et al. and Segura et al.) generally present centimeter-level accuracy but all of them rely in complex and expensive hardware that is not compatible with conventional smartphones.

The NOKIA HAIP and Ubisense systems are both used in applications where increased accuracy and high position estimation rate is needed. A major drawback of these two systems is the high infrastructure costs.

The classification used for comparison of the technology and ranging method was based in the taxonomy introduced in Section 2.14, cf. Figure 2.14. The accuracy metric is presented in three distinct formats depending on the metric used for evaluation by the authors: (1) as an interval, (2) using x (p%), which means that in p percent of the cases the accuracy is less than x or (3) as x ($\pm s$), where x is the accuracy and s is the standard deviation. The system scalability in terms of simultaneous users is divided in two degrees: Limited or High. The bandwidth used for the signal design is presented as an absolute value or an interval depending on the information available in the references. The update rate is presented as the maximum value in Hz and the system cost is divided in two categories: Inexpensive or Expensive.

Name	Category	Year	Technology	Ranging Method	Reported Accuracy	Simultaneous Users	Bandwidth	Update Rate	Cost	Smartphone Compatibility
Guoguo [30]	Acoustic	2013	Audio/US	MT-TDoA	6-25 cm	High	17-20 kHz	$\leq$ 2 Hz	Inexpensive	yes
LOSNUS [31]	Acoustic	2010	Audio/US	MT-TDoA	1 cm	High	35-65 kHz	$\leq$ 10 Hz	Expensive	no
3D Locus [32]	Acoustic	2009	Audio/US	RToF	0.5 cm (90%)	Limited	<25 kHz	10 Hz	Inexpensive	no
Parrot [33]	Acoustic	2006	US+auxRF	ST-TDoA	2 cm ($\pm$ 4)	High	1 kHz	$\leq$5 Hz	Expensive	no
Cricket [35]	Acoustic	2005	US+auxRF	ST-TDoA	1-2 cm	High	1 kHz	1 Hz	Inexpensive	no
Beep [36]	Acoustic	2005	Audio	ToA	61 cm(97%)	Limited	4 kHz	-	Inexpensive	yes
Nokia HAIP [37]	RF	2012	Bluetooth	AoA	30-100 cm	High	1 MHz	0.1-40 Hz	Expensive	yes
UBISENSE [38]	RF	2011	UWB	AoA/MT-TDoA	<15 cm	Limited	-	-	Expensive	no
Pietrzyk et al. [42]	RF	2010	UWB	ToA	1-2 cm	Limited	-	-	Expensive	no
Segura et al. [43]	RF	2010	UWB	MT-TDoA	<20 cm	Limited	-	-	Expensive	no
COMPASS [40]	RF	2005	WiFi/FP	RSS	<2 m (70%)	High	-	-	Inexpensive	yes
RADAR [39]	RF	2000	WiFi/FP	RSS	<4.7 m (75%)	High	-	0.25 Hz	Inexpensive	yes

Table 2.2: Performance of selected centimetre-level positioning systems.

2.3 Positioning Process and Related Taxonomy

In [44], Amundson proposes a conceptual organization for the positioning process based on three main stages (Figure 2.13): 1) Coordination, 2) Measurement and 3) Position Estimation.

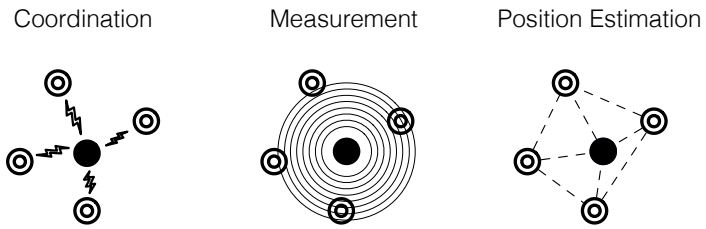


Figure 2.13: Positioning Process. Adapted from [44].

The coordination stage is normally related with synchronization. Synchronization is important in the measurement stage and is typically used to notify all the involved nodes that a measure will be taken. Depending on the method used in positioning, synchronization mechanisms between nodes should be taken into account.

The measurement stage is typically related with a signal transmission and is normally used to measure distances or angles between nodes. These geometric relations between nodes can be obtained using different signal types (e.g. RF or acoustic) and different measurement techniques (e.g. RSS-based, Time-based and Angle-based), as described in Figure 2.14 [45]. RSS-based systems are known as a less accurate approach that uses the RF Received Signal Strength (RSS) measurements to obtain distance estimates between transmitters and receivers based on location-dependent characteristics. RSS measurements are commonly used in scene analysis methods (fingerprinting) but can also be used to estimate distances (using a path loss model to convert RSS values into distances) or angles (by using directional antenas) which also enables the use of RSS measurements with trilateration or triangulation positioning algorithms.

The position estimation stage uses measurements obtained in the measurement stage to compute the target node position. There are several approaches for position estimation, but due to the scope of this work, we will only focus on the most accurate for indoor applications, namely Scene Analysis (fingerprinting) and Geometry-based methods.

The signal type used, along with the selected measurement principle, heavily affects the position estimation accuracy. For example, the often low accuracy that RSS-based systems provide, limits their use with geometric-based positioning approaches, such as trilateration. A common method is to use RSS measurements with a scene analysis approach to obtain position estimates with accuracies that can reach the meter level. This approach has several limitations that will be addressed in section 2.3.1.1. On the other hand, accurate indoor positioning schemas rely on geometric relations between the beacons and target nodes. These relations are normally obtained by measuring the

propagation time using Time of Arrival (ToA), Time Difference of Arrival (TDoA) measurements or by measuring the Angle of Arrival (AoA) of the incoming signals at the target node.

Figure 2.14 describes a taxonomy for the classification of the most relevant measurement techniques and positioning approaches under the scope of this work. In the following sections they are described in detail with respect to their working principles, major advantages and limitations.

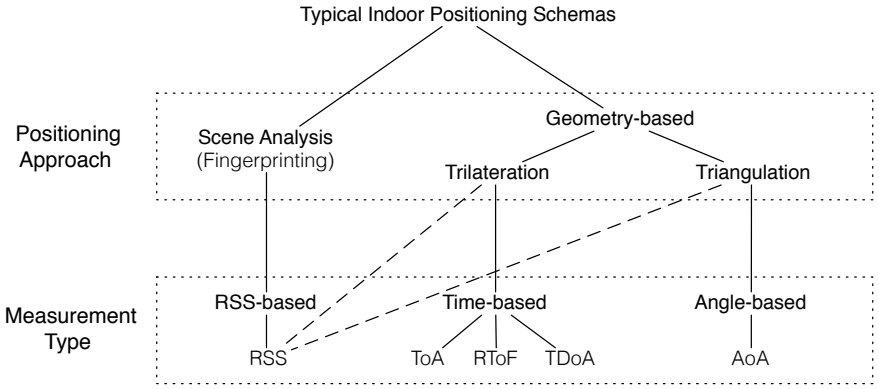


Figure 2.14: Classification of typical indoor positioning schemas used for tracking applications.

2.3.1 Measurement Type

In this section, only range-based techniques will be addressed. All the techniques presented in Figure 2.14 are normally used in indoor user tracking applications. The three main measurement types we focus are : 1) RSS-based, 2) Time-based and 3) Angle-based.

2.3.1.1 RSS-based

Received Signal Strength (RSS) is a RF power measurement obtained at the receiver used to estimate the distance between a transmitter and a receiver. This method uses either a theoretical or empirical model to translate signal strength into distances [46]. Knowing the transmitted power, and the path loss model, the receiver estimates the distance to the transmitter using the power strength of the received RF signal. Theoretically this attenuation can be simply derived, but in the real world, when nodes are mobile, fading and multipath problems occur making RSS measurements less effective. To solve this problem, statistical methods based in multiple measurements are normally used [9].

For an ideal environment, a simple path loss propagation model can be used to convert the RSS power measurements into distance estimates, where the RSS values are inversely proportional to d^{η} (where d is the distance between transmitter and receiver and η refers to the path loss exponent [47]). The path loss exponent is an environment-dependent parameter that represents the signal attenuation for a given environment. Generally, the path loss model is given by [48]:

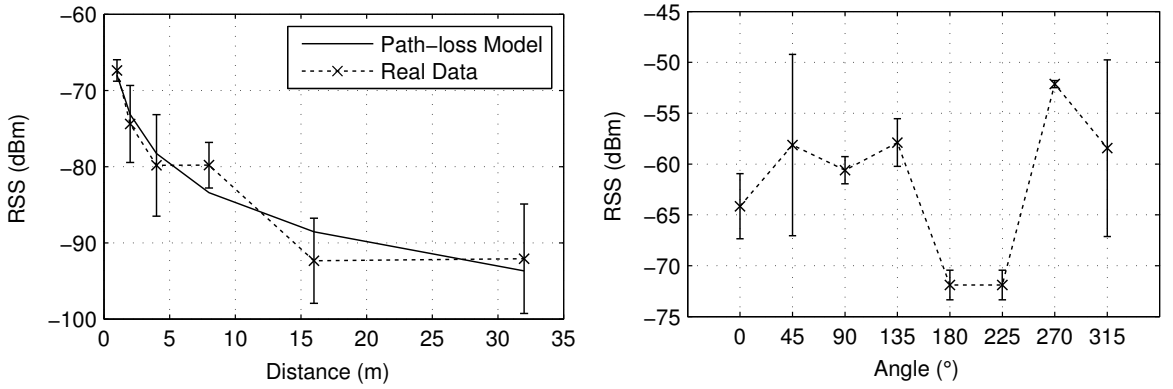
$$\text{RSS} = P_t \times K \times \left(\frac{d_0}{d}\right)^{\eta} \tag{2.1}$$

where P_t is the transmitted power, K is a constant related with the transmitter/receiver gains, the carrier frequency and the path loss at a reference distance d_0 (normally at 1 m). A common RSS-distance mapping model is the log-normal distance path loss propagation model [48]:

$$\text{RSS}(d)[dB] \quad = \quad \text{RSS}(d_0) - 10\eta \log(d/d_0) + X_\sigma[dB] \tag{2.2}$$

where RSS(d) represents the received signal strength, RSS(d_0) is the received signal strength at a reference distance d_0, η represents the path loss exponent and X is a zero-mean Gaussian random variable with a variance σ^2, which is referred to as the shadowing variance [8].

When used indoors, and due to multipath propagation (i.e. reflection, refraction, diffraction and scattering) of the RF signals, the mapping of RSS to distance based on the path loss model introduced in Equation 2.2 is demanding, due to the high variability that RSS measurements present. The two major problems affecting indoor RSS estimation are the multipath and fading phenomena of the indoor RF channel and the heavy dependence of the measurements on the radiation pattern of both transmitter and receiver antennas.



(a) (b)

Figure 2.15: (a) Log-normal distance path loss propagation model (Equation 2.2) with RSS(d_0) = -68 dB and $\eta = 0.87$ and real data obtained indoors using Bluetooth Low Energy RSS measurements. (b) RSS measurement dependence on the orientation of the transmitter antenas.

Figure 2.15a presents the measured RSS variability obtained indoors using a Bluetooth Low Energy Sensor Tag (from Texas Instruments) acting as a transmitter in LoS with an iPhone 4S

working as a receiver. From the data it is possible to observe a high variability (high standard deviation) in most measurements. This high variability results in unreliable distance estimates that are not suitable for high accuracy indoor positioning systems. Additionally, Figure 2.15b also presents the high RSS dependence with the angle between transmitter and receiver. This plot was obtained when the transmitter and the receiver were in LoS at a distance of 1 m and the transmitter was rotated around itself in intervals of 45° for a set of one hundred measurements in each rotation. The resulting RSS varies up to 20 dB and the standard deviation up to 10 dB. These observations can be justified by the conjugation of the transmitter and receiver radiation patterns.

RSS has been used for range estimation in [49, 50] and in the following indoor user tracking systems [39, 40, 51] with reported accuracies below 3 m.

The main advantages and limitations of RSS-based techniques are:

Advantages:

- Low cost and wide availability, e.g. WiFi access points are common and already deployed in buildings.
- Range estimates are clock-free (no need of synchronization mechanisms).
- Can be obtained using simple algorithms, cf. [50].

Limitations:

- Low accuracy (meter-level), due to high variability of the relevant parameters of the path loss model in indoor environments.
- RF transceivers must be calibrated, so that one signal strength cannot result in different RSS values on different devices.

2.3.1.2 Time-based

These methods estimate the distance between nodes by measuring the signal propagation time (RF or acoustic) between a transmitter and a receiver, using one of these three main approaches: Time-of-Arrival (ToA), Round-trip Time of Flight (RToF) and Time Difference of Arrival (TDoA). Next, these three methods are discussed in more detail.

Time of Arrival (ToA)

This method consists in calculating the one way propagation time (RF or acoustic) between two synchronized nodes. The distance between these nodes can be estimated if the wave propagation speed in the medium is known. In this technique, receiver and transmitter nodes, must be synchronized to compute the ToA at the receiver, cf. Figure 2.16.

By measuring the propagation time, and knowing the propagation speed, a distance can be directly estimated. If no errors are considered in the synchronization process the distance is given

by

$$d = v \times (t_1 - t_0), \tag{2.3}$$

where v represents the speed of propagation of the signal (RF or acoustic) t_0 and t_1 are the transmission and reception instants, respectively. This schema requires the receiver to know the accurate instant of transmission (t_0), which can be achieved using a synchronized system, i.e. the transmitter and the receiver have a common clock. This type of approach is normally used when accurate ranging is needed and when a synchronization strategy between all intervenient nodes is available. Moreover, this time-based method presents low communication and computation overheads.

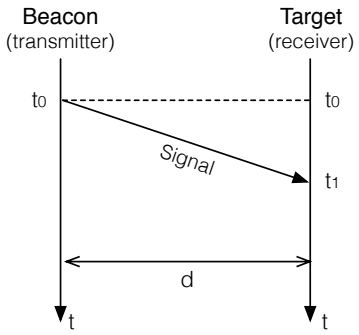


Figure 2.16: Time of Arrival (ToA) method.

ToA has been used for range estimation in [42, 43, 25, 30] with reported accuracies below 25 cm. The main advantages and limitations of the ToA technique are:

Advantages:

- Very accurate position estimates can be obtained with an optimal beacon geometry.
- Beacons and target nodes can work as transmitters or receivers. There is no need to use transceivers in all intervenient nodes.

Limitation:

- Effective synchronization between all intervenient nodes is mandatory. If RF signals are used for accurate radio ranging, the synchronization should be guaranteed in the nanosecond order which results in expensive and complex hardware design.

  On the other hand, if acoustic signals are considered, synchronization can be easily achieved, by using the time difference of arrival of a RF sync pulse and an acoustic pulse transmitted at the same time. In this document we refer to this method as single-transmitter TDoA and it will be introduced later in this chapter.

Round-trip Time of Flight (RToF)

This method is a variation of ToA and it is normally used to avoid the synchronization requirement between transmitter and receiver imposed by ToA systems. It consists in the measurement of the round-trip ToF by the transmitter using its clock to compute the propagation round-trip delay, thus avoiding synchronization between transmitter and receiver, cf. Figure 2.17.

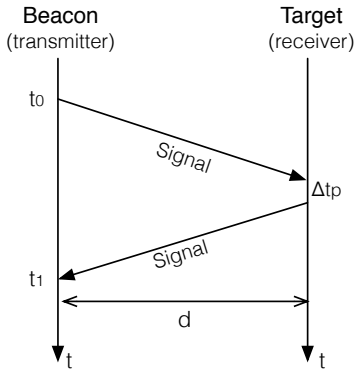


Figure 2.17: Round-trip Time of Flight (RToF) method.

Note that there is a processing delay (Δt_p) in the receiver that should be estimated and removed in the final round-trip ToF measurement by using a preliminary calibration process or by sending its value to the transmitter for each distance measurement. For this approach, the distance between the transmitter and the receiver is given by Equation 2.4.

$$d = v \times \frac{(t_1 - t_0) - \Delta t_p}{2} \tag{2.4}$$

where v represents the speed of propagation of the signal (RF or acoustic) and $(t_1 - t_0) - \Delta t_p$ represents the round-trip ToF.

This type of approach is used when accurate ranging is needed and when a synchronization strategy between intervenient nodes is not possible. This time-based method presents low communication and computation overheads. RToF is used in [32] with reported accuracies below 2 cm.

The main advantages and limitations of the RToF technique are:

Advantages:

- Syncronization between nodes is not mandatory.
- Very accurate position estimates can be obtained with an optimal beacon geometry.
- When compared with ToA systems, it can be less expensive than a synchronized system.

Limitations:

- All nodes in the system must be transceivers.
- Computational delays due to roundtrip and interpretation of signals.
- Higher channel occupation, which results in a slower measurement rate.

Time Difference of Arrival (TDoA)

In the literature there are two different definitions for the term Time Difference of Arrival (TDoA). One definition will be referred as Single-Transmitter Time Difference of Arrival (ST-TDoA) and uses the time difference obtained when two transmission mediums with different propagation speeds are used to obtain an implicit synchronization between a transmitter and a receiver. The other definition originates from geolocation systems [52] and will be referred as Multiple-Transmitter Time Difference of Arrival (MT-TDoA). It consists in the computation of the time difference of arrival when two or more synchronized transmitters at different distances from the receiver transmit simultaneously.

Single-Transmitter Time Difference of Arrival (ST-TDoA): In this approach the transmitter node simultaneously sends two signals with different propagation speeds, e.g. a RF pulse and an acoustic pulse, cf. Figure 2.18.

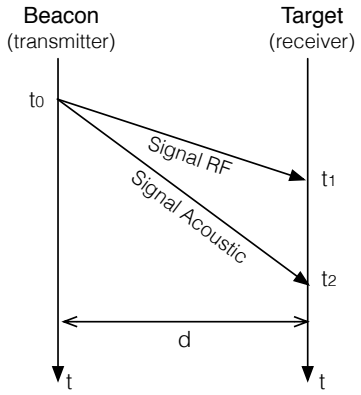


Figure 2.18: Single-Transmitter Time Difference of Arrival (ST-TDoA) method.

As the propagation speed of these signals differ in several orders of magnitude $\approx 10^6$, the receiver uses the arrival time of the radio signal (arrives first) as a time reference to calculate the delay between both signals. This can be seen as an implicit synchronization strategy used to measure the arrival time of the acoustic signal at the receiver. On the other hand, this technique implies the use of additional hardware to deal with the two distinct types of signal used, which is a great disadvantage. Figure 2.18 presents an example of this technique using RF and acoustic pulses. The distance d between the transmitter and the receiver is computed using Equation 2.5:

$$d = \frac{v_r \times v_a \times (t_2 - t_1)}{v_r - v_a} \tag{2.5}$$

where v_r and v_a represent the propagation speed of both radio and acoustic signals and t_1 and t_2 are its reception instants at the receiver, respectively. ST-TDoA has been used in [33] and [35] with reported accuracies below 2 cm.

The main advantages and limitations of the ST-TDoA technique are:

Advantages:

- Very accurate position estimates can be obtained for an optimal beacon geometry.
- Synchronization between nodes is not mandatory.

Limitation:

- Each node needs two different types of transceivers which results in complex and costly hardware.

Multiple-Transmitter Time Difference of Arrival (MT-TDoA): This approach is based on the propagation of only one type of signal (RF or acoustic) and needs at least two transmitters to estimate a time difference of arrival value at the receiver, cf. Figure 2.19. If the transmitters are at different distances from the receiver and transmit synchronized pulses, the receiver will receive the pulses at different instants t_1 and t_2.

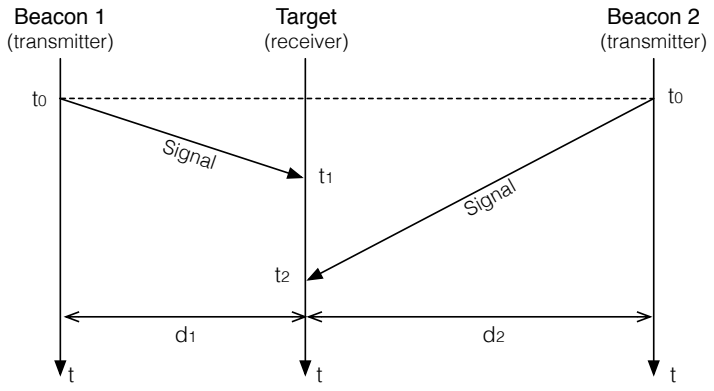


Figure 2.19: Multiple-Transmitter Time Difference of Arrival (MT-TDoA) method.

The MT-TDoA obtained $(t_2 - t_1)$ results from the distance differences $(d_2 - d_1)$ of the transmitters to the receiver. In this case the transmitter position is given by a hyperbola whose generic equation is given by:

$$D_{i,j} = d_i - d_j = v \times (t_i - t_j) \tag{2.6}$$

where d_i is the distance between transmitter i and the receiver node

$$d_i = \sqrt{(x_i - x_t)^2 + (y_i - y_t)^2 + (z_i - z_t)^2}, \tag{2.7}$$

were (x_i, y_i, z_i) represent the 3D coordinates of a fixed transmitter i and (x_t, y_t, z_t) represent the coordinates of the target node. If the two transmitters positions are known, and the TDoA at the receiver is obtained, the locus of the receiver is restricted to half of a two-sheeted hyperboloid. To find the 2D analytical solution to this problem, an additional TDoA estimate is required, i.e. an additional hyperboloid. The 2D locus of the receiver will be the intersection of these two hyperboloids. In this approach, the receiver does not need to be synchronized with the transmitters since only the time instants of arrival are need to compute the TDoA values. This method is also known in the literature as hyperbolic positioning.

MT-TDoA has been used in [31, 43, 38] with reported accuracies below 20 cm. The main advantages and limitations of the MT-TDoA technique are:

Advantages:

- Synchronization is mandatory only between transmitters, because the clock bias can be canceled out by taking the difference of the times of arrival of signals from distinct transmitters.
- Less expensive than ToA systems.

Limitations:

- The positioning accuracy is worse than in ToA systems with the same system geometry.
- The clock bias is involved in the range equations in a linear manner, which makes the system geometry design of a MT-TDoA schema more complicated than a ToA schema, since a rank defect situation can happen in more geometries with a MT-TDoA system [9].

2.3.1.3 Angle-based

Angle-based methods mainly use the AoA technique to estimate the angle at which signals are received and simple geometric relations to compute the azimuth to neighbor nodes with respect to some reference direction [53]. This reference direction is also known as orientation and may be defined as absolute if it refers to the real North direction, cf. Figure 2.20a. Otherwise, it can be defined as relative when a relative reference with respect to the North direction is known in advance, cf. Figure 2.20b.

By knowing the relative or absolute orientation of a receiver, only two transmitters acting as beacons are needed to estimate the position of the receiver. Nevertheless, if the orientation is

unknown, three non-colinear transmitters acting as beacons are needed to estimate the position of the receiver, cf. Figure 2.20c.

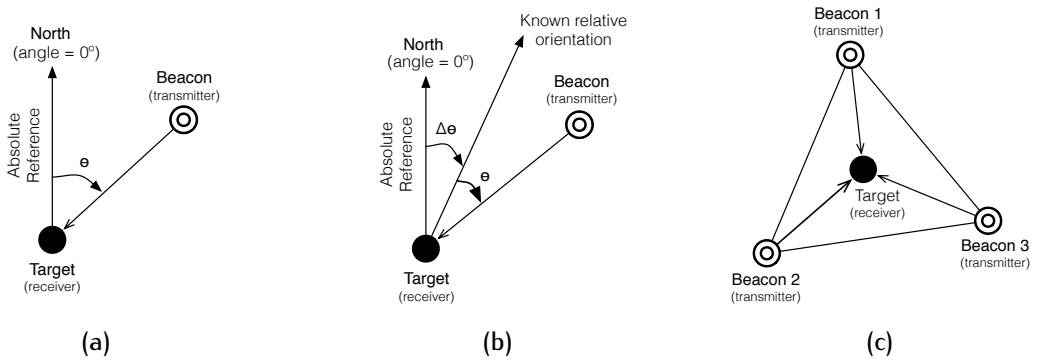


Figure 2.20: AoA with absolute, relative and unknown orientation. Adapted from [48]. (a) AoA with absolute orientation; (b) AoA with relative orientation and (c) AoA with unknown orientation.

Geometrically, the estimated angle $\hat{\theta}$ between the target node with coordinates (x_t, y_t) and a beacon acting as a transmitter with known coordinates (x_i, y_i) is determined by [48]:

$$\hat{\theta} = \theta + n_i \tag{2.8}$$

were

$$\theta = \arctan\left(\frac{x_t - x_i}{y_t - y_i}\right) \tag{2.9}$$

and n_i is zero-mean Additive White Gaussian Noise (AWGN) with variance σ^2.

The AoA technique is commonly used in two approaches: (1) using an antenna array topology (Figure 2.21a), or (2) based on the RSS ratio between two or more directional antennas located on the node itself (Figure 2.21b) [27].

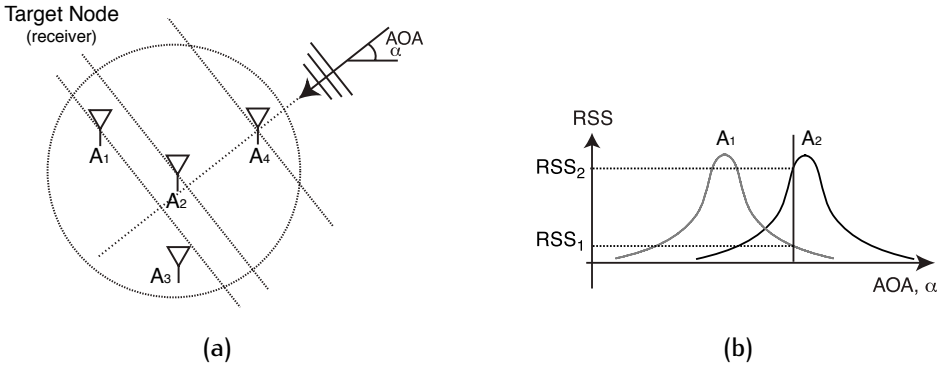


Figure 2.21: AoA typical approaches. (a) AoA using an antenna array; (b) AoA using multiple directional antennas to measure RSS. A_i represents the ith antenna. Adapted from [27].

AoA schema, when compared with the time-based methods, requires an antenna array (instead of a single antenna) at the receiver to estimate the AoA of the received signal. However, we should note that the AoA schema also has some common features with the time-based methods, since the AoA of the received signal regarding a certain transmitter is usually estimated based on the time difference of arrival of the received signal at the antenna array elements in the receiver. AoA has also been used for position and orientation estimation in an ad-hoc network in [54] and for positioning in [37, 38] with reported accuracies below 30 cm.

The main advantages of the AOA-based methods are:

Advantages:

- Time synchronization between nodes is not required since the AoA is obtained using the (pseudo) range differences of multiple antennas at the same target node.
- Fewer transmitters (when compared RSS and Time-based schemas) are required to estimate the final position.

Limitations:

- Requires complex, oversized and expensive hardware (antenna array) which compromises important hardware design requirements such as low-cost and low-power.
- The computation requirements of AoA estimation is complex and heavily affected by noise, scattering, shadowing and multipath wave propagation, which imposes challenges on accurate AoA estimation.

2.3.2 Positioning Approach

In this section, we introduce and discuss the most important positioning approaches used indoors that deliver increased accuracies. As presented in Figure 2.14, the two main positioning approaches we focus are the Scene Analisys and Geometry-based methods.

2.3.2.1 Scene Analysis

Typically, scene analysis methods use RSS measurements; they can also be used with acoustic signals or images [8]. Scene analysis, also known in literature as fingerprinting (FP), is commonly divided in two main stages: the training stage and the positioning stage. The training stage consists in a prior off-line calibration process to build a map of "fingerprints" that are empirically obtained by taking exhaustive RSS measurements inside a building. In the positioning stage, a position estimate is obtained by matching on-line measurements with the best position in a database that approaches the pre-computed "fingerprint" [47]. The most common algorithm to estimate the position of a target node computes the euclidean distance between the measured RSS vector and each fingerprint in the database [8] and then the coordinates associated with the fingerprint that provides the smallest euclidean distance are selected as an estimate of the position.

If maps are included, accuracy can be improved due to the inclusion of several restrictions that buildings impose on the decisions of the positioning stage. There are techniques used to reduced the human effort needed in the training stage, by including an automated mapping process, usually called radio maps. Scene analysis has been used in the indoor positioning systems [55, 56] with accuracies below 2 m.

The main advantages and limitations of the scene analysis methods are:

Advantages:

- Low cost commecial-of-the-shelf (COTS) hardware. Method commonly used with RSS RF measurements.
- High availability. If the IEEE 802.11 standard is adopted, devices such as WiFi routers, smartphones, laptop computers, tablets or generic WiFi nodes can be used.

Limitations:

- The profiling operation is too complex, increasing with the area of the explored environment and the number of devices to be deployed in the training phase.
- Performance is sensitive to indoor environmental changes. Movement of persons/objects, will compromise the validity of the static mapping previously obtained in the training phase, which directly results in loss of accuracy.
- The final positioning accuracy depends on the number and the distribution of access points. A greater number of access points may increase the positioning accuracy, but this requires a more complex training stage and increases the processing time in the positioning stage.

2.3.2.2 Geometry-based

In this type of systems, geometric relations (distances or angles) between the network nodes are used to obtain a position estimate. If distance measurements are obtained it is possible to estimate the position of the target node using the trilateration approach. Alternalively, if angle measurements are available it is possible to find the target node position using the triangulation approach.

Trilateration

Trilateration estimation can be used to find the target node position if three or more non-colinear beacons are used. This method uses the inter-node distances to obtain a position estimate of the target node. Figure 2.22 illustrates a particular case of trilateration when three beacons are considered. In this example, localization is performed by calculating the intersection of the three circles centered at the beacon coordinates and radius d_i. Each range measurement is here represented as the radius of each circle. Note that, if the measurements contain errors, the intersection of the three circles is an area instead of a single point [53]. This situation leads to an optimization problem

that can be solved using an estimation method to minimize the error in positioning with relation to all the previous measured distances.

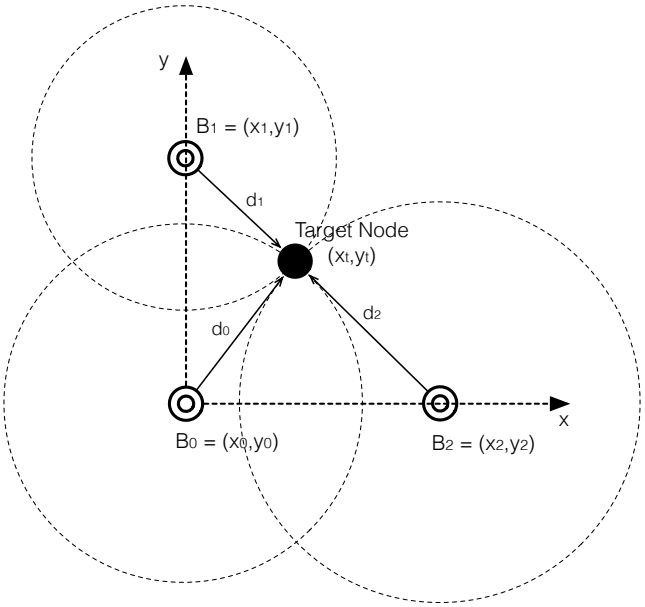


Figure 2.22: Trilateration example with three beacons.

Problem Statement : For the 2D trilateration example presented in Figure 2.22, a system of non-linear equations can be obtained, to solve for (x_t,y_t). For each beacon a circle with radius d_i is presented. This radius is the distance obtained (normally using ToA range estimates) using the following equation,

$$d_i = v \times (t_i - t_0) \tag{2.10}$$

where v is the wave propagation speed, t_i is the time instant of arrival and t_0 is the transmission time instant.

For simplification we consider beacon B_0 at the origin, i.e. $B_0 = (0, 0)$. In this case, distance estimates d_i can be used to estimate the target node position (x_t, y_t) by solving the following system of equations:

$$d_0^2 = x_t^2 + y_t^2 \tag{2.11}$$

$$d_1^2 = (x_1 - x_t)^2 + (y_1 - y_t)^2 \tag{2.12}$$

$$d_2^2 = (x_2 - x_t)^2 + (y_2 - y_t)^2 \tag{2.13}$$

Analytical Solution : The overdetermined non-linear system of equations can be solved analytically to obtain the position coordinates of the target node (x_t,y_t) resulting in,

$$x_t = \frac{A_0 Y_{11} + A_1 Y_{02} + A_2 Y_{10}}{2(x_0 Y_{21} + x_1 Y_{02} + x_2 Y_{10})} \tag{2.14}$$

$$y_t = \frac{A_0 X_{21} + A_1 X_{02} + A_2 X_{10}}{2(y_0 X_{21} + y_1 X_{02} + y_2 X_{10})} \tag{2.15}$$

were

$$A_k = x_k^2 + y_k^2 - d_k^2 \quad \text{with} \quad 0 \leq k < 3 \tag{2.16}$$

$$X_{ij} = (x_i - x_j) \quad \text{with} \quad 0 \leq i < 3 \quad \text{and} \quad 0 \leq j < 3 \tag{2.17}$$

$$Y_{ij} = (y_i - y_j) \quad \text{with} \quad 0 \leq i < 3 \quad \text{and} \quad 0 \leq j < 3 \tag{2.18}$$

Linear Least Squares (LLS) Approximation : Given that distance measurements (d_i) are always noisy, the position estimation can be seen as an optimization problem. The Squared Differences Method is a simple linearization method where the solution for the set of non-linear equations presented by Equations 2.11, 2.12 and 2.13 can be obtained by subtracting Equation 2.11 from Equations 2.12 and 2.13 as

$$d_1^2 - d_0^2 = x_1^2 - 2x_1 x_t + y_1^2 - 2y_1 y_t \tag{2.19}$$

$$d_2^2 - d_0^2 = x_2^2 - 2x_2 x_t + y_2^2 - 2y_2 y_t \tag{2.20}$$

Rearranging the terms, the above two equations can be written in matricial form as

$$\begin{bmatrix} x_1 & y_1 \\ x_2 & y_2 \end{bmatrix} \begin{bmatrix} x_t \\ y_t \end{bmatrix} = \frac{1}{2} \begin{bmatrix} K_1^2 - d_1^2 + d_0^2 \\ K_2^2 - d_2^2 + d_0^2 \end{bmatrix} \tag{2.21}$$

where

$$K_i^2 = x_i^2 + y_i^2. \tag{2.22}$$

The linearized system of equations presented in 2.21 can be rewritten as

$$\mathbf{H}\hat{\mathbf{x}} = \mathbf{b} \tag{2.23}$$

with

$$\mathbf{H} = \begin{bmatrix} x_1 & y_1 \\ x_2 & y_2 \end{bmatrix} \quad , \quad \hat{\mathbf{x}} = \begin{bmatrix} x_t \\ y_t \end{bmatrix} \quad , \quad \mathbf{b} = \frac{1}{2} \begin{bmatrix} K_1^2 - d_1^2 + d_0^2 \\ K_2^2 - d_2^2 + d_0^2 \end{bmatrix}. \tag{2.24}$$

The approximate solution $\hat{\mathbf{x}}$ is given by

$$\hat{\mathbf{x}} = (\mathbf{H}^T\mathbf{H})^{-1}\mathbf{H}^T\mathbf{b} \tag{2.25}$$

where $\mathbf{H}^T$ is the transposed $\mathbf{H}$ matrix and $(\mathbf{H}^T\mathbf{H})^{-1}$ is the pseudo-inverse operator.

Triangulation

Angulation estimation is a trigonometric approach used for determination of an unknown position using two angles and the distance between them. If two angles between beacons and the target node can be measured (normally based in the AoA technique) angulation can be used to perform localization by applying simple trigonometry methods [53].

Angulation can be computed with at least two beacons. The position will be calculated as the third point in a triangle of two known angles, and the distance (d_i) between the two anchor nodes [44]. If three beacons are used, this technique is normally called triangulation and the position of the target can be computed using the simple geometric relations of two triangles, cf. Figure 2.23. In noisy measurements, as in lateration, the intersection will not be a point, but instead, an area that defines the region where the node is supposed to be. In this case, we have also an optimization problem that can be solved using an estimation method to minimize the positioning error with relation to all the measured angles, cf. [57].

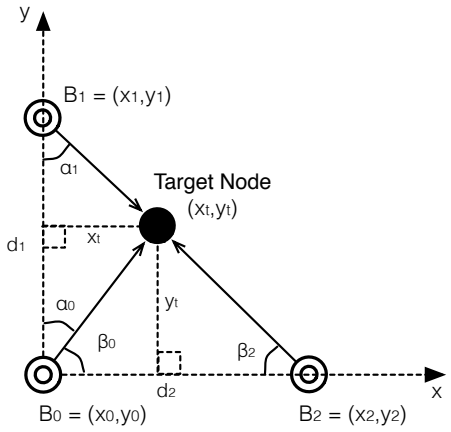


Figure 2.23: Triangulation example with three beacons.

Problem Statement : For the 2D triangulation example presented in Figure 2.23, the target node position (x_t, y_t) can be determined by using the combination of B_0 and B_2 to find x_t, and the combination of B_0 and B_1 to find y_t. Note that, beacons B_0 and B_2 form the baseline of the x-axis and beacons B_0 and B_1 are used to form the baseline of the y-axis.

Analytical Solution : The resulting system of equations can be solved analytically using trigonometric operations to obtain the position coordinates of the target node (x_t,y_t) resulting in,

$$x_t = \frac{d_1 \sin(\alpha_0) \sin(\alpha_1)}{\sin(\alpha_0 + \alpha_1)} \tag{2.26}$$

$$y_t = \frac{d_2 \sin(\beta_0) \sin(\beta_2)}{\sin(\beta_0 + \beta_2)} \tag{2.27}$$

2.4 Chapter Summary

In this chapter we discussed the technological diversity in indoor positioning systems with five categories defined: Optical-based, Mechanical-based, Electromagnetic-based, Acoustic-based and RF-based systems. Additionally, a comparison and discussion of the studied centimeter-level positioning systems is undertaken based in the two most important categories previously defined: Acoustic-based systems (i.e. Guoguo, LOSNUS, 3D-LOCUS, Parrot, Cricket and Beep) and RF-based systems (i.e. NOKIA HAIP, Ubisense UWB, Pietrzyk et al. UWB Experimental Plataform, Segura et al. UWB, COMPASS and RADAR). Then, an overview of the positioning process was introduced, followed by a proposed taxonomy that relates the most important techniques and methods used in indoor positioning systems; in relation to the measurement type (RSS-based, Time-based and Angle-based) used in the ranging process, and in the positioning approach (Scene Analysis and Geometry-based) used for position estimation.

CHAPTER 3

Non-Invasive Acoustic Ranging

Sound can be seen as a mechanical disturbance of the medium, which may be solid, liquid or gas (e.g. air). The indoor acoustic environment has an important effect in the way that sound waves propagate in space. Figure 3.1 illustrates the generic point-to-point acoustic ranging case and identifies all the relevant topics that should be taken into account in the development of an indoor non-invasive acoustic ranging system.

This type of systems is mainly affected by the indoor environment (sound propagation issues, room acoustics and background noise), the transducer (transmitter/receiver) characteristics and the perceptual impact of sound on humans. Measuring distances based on acoustic signals has some advantages [58] when compared to traditional RF methods. Sound propagation is isotropic and has predictable signal attenuation leading to accurate ranging results [45]. In [59, 24, 60, 61] acoustic signals have been used for positioning and the results were reported with accuracy in the centimeter-level order.

In this chapter we introduce the theory related to the most important topics of the ranging process to obtain a simplified indoor acoustic model, cf. Section 3.1.1, based on the following sub-models: (a) Sound Propagation Model, (b) Room Model, (c) Noise Model and (d) Transducer Model.

As we are interested in the design of non-invasive pulses for ranging, it is important to understand the physical limitations of the human auditory system and how sound loudness is perceived by humans. In Section 3.2.2 we address these topics to find a way to reduce the perceptual impact of pulses with spectral content in the audio band.

In Section 3.3 we characterize the audio input bandwidth for several commercially available smartphones and in Section 3.4 we introduce the pulse detection theory for the matched filter case.

In the end of this chapter, we address the non-invasive pulse design approaches (Section 3.5), followed by the description of the perceptual evaluation procedure used in the evaluation with subjects, cf. Section 3.6.

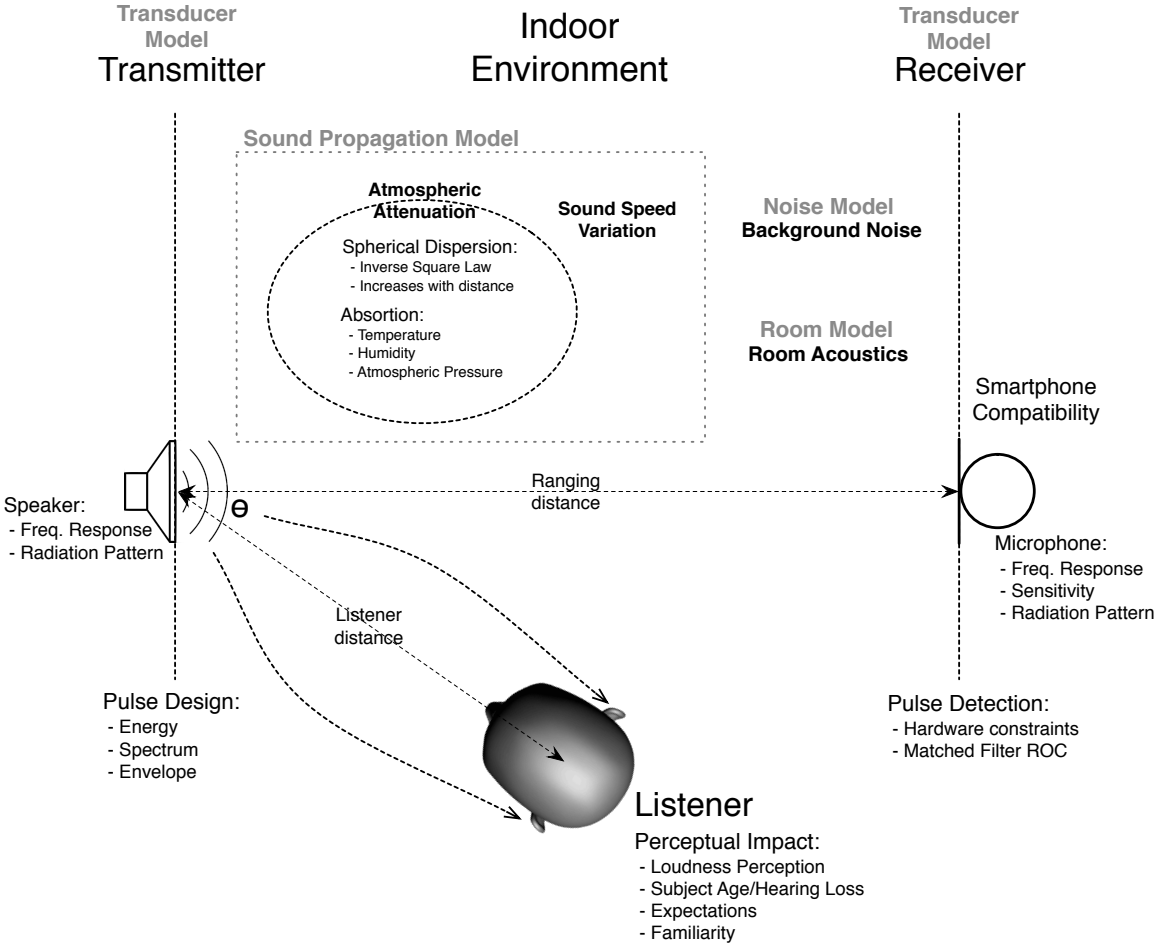


Figure 3.1: Representation of the non-invasive acoustic ranging and relevant topics.

3.1 Indoor Acoustics Modeling

Ultrasonic signals, for example at 40 kHz, are negatively impacted by severe attenuation for distances higher than 10 meters [62, 63], which is a major limitation for ranging applications. On the other hand, audible sound signals can be very effective, because of the wide diversity of wavelengths that can be used [62].

Acoustic propagation suffers from absorption, when traveling in the air due to different physical processes (such as variations in temperature and humidity) and from dispersion, which is observed by

the geometric decrease in intensity with increasing distance. Air absorption depends on frequency. It is quite low for audible frequencies, but it becomes extremely large as the frequency increases in the ultrasonic band. Using acoustic signals at lower frequencies, such as in the audible range, it is possible to measure longer distances, thus circumventing the high attenuation that ultrasounds suffer. Existing audible ranging systems have shown that it is possible to measure distances as large as 100 m, under ideal weather conditions [62].

A common model for sound propagation in space is the free field wherein it is considered that the sound propagation is isotropic, free from boundaries and is undisturbed by other sources of sound (including noise) [64]. This simple approach will be used to evaluate the sound speed during propagation and to evaluate the attenuation of sound waves due to spherical spreading (also known as the inverse-square-law) and atmospheric absorption which is only dependent on physical variables, such as temperature, humidity, air composition and pressure rather than the surrounding acoustic environment, such as walls or objects. Additionally, a simplified room acoustic model based on the free field model with added boundaries will be used to evaluate the impact of multipath propagation (which includes early reflections and reverberation) in an acoustic ranging system. A simple background acoustic noise model will be defined based on the inclusion of Additive White Gaussian Noise (AWGN) and a piston-based transducer model will be proposed.

3.1.1 Sound Propagation Model

Sound speed changes and sound attenuation due to atmospheric absorption and spherical dispersion have significant impact on sound propagation and will be discussed in detail in the following subsections for the free field acoustic environment model.

3.1.1.1 Speed of sound

The sound speed propagation can be modeled using the adiabatic gas law equation because the mechanical disturbance moves so quickly that there is no time for heat to transfer from the compressions or rarefactions [65]. If we consider the generic adiabatic gas law,

$$PV^{\gamma} = constant \tag{3.1}$$

where P is the air pressure of the gas (in $N{\cdot}m^{-2}$), V is the volume of the gas (in m^3) and γ is a constant which depends on the gas (1.4 for air). From Equation 3.1 we can obtain a relationship between the pressure and volume of a gas. This relation can be used to determine the elastic modulus of the air, which is also known as the Young's modulus air equivalent,

$$E = \gamma P \tag{3.2}$$

The ideal air density is obtained by,

$$\rho = \frac{m}{V} = \frac{PM}{RT} \tag{3.3}$$

where m represents the mass of the gas (in kg), M is the molecular mass of the gas (in kg·mole^{-1}), R is the gas constant (8.31 J·K^{-1}·mole^{-1}) and T the absolute temperature in K. From Equations 3.2 and 3.3 we can obtain the equation for the speed of sound in air (v_{air}),

$$v_{\text{air}} = \sqrt{\frac{E_{\text{air}}}{\rho}} = \sqrt{\frac{\gamma P}{(\frac{PM}{RT})}} = \sqrt{\frac{\gamma RT}{M}} \tag{3.4}$$

Using Equation 3.4 and considering a simplified air composition [65] of 21% oxygen (O_2), 78% nitrogen (N_2), 1% argon (Ar) the resulting molecular weight of air is given by,

$$M = 21\% \times 16 \times 2 + 78\% \times 14 \times 2 + 1\% \times 18 M = 2.87 \times 10^{-2} \text{ kg·mole}^{-1} \tag{3.5}$$

we can obtain the speed of sound, that is only dependent on the square root of the absolute temperature:

$$v_{\text{air}} = \sqrt{\frac{1.4 \times 8.31 \times T}{2.87 \times 10^{-2}}} = 20.1\sqrt{T} \tag{3.6}$$

Two important observations can be made by looking at Equation 3.6. The first is that the speed of sound in air is not affected by pressure and the second is that the speed of sound is strongly related to the absolute temperature and the molecular weight of the gas.

Considering T=18°C, the speed of sound can be obtained directly from Equation 3.6, resulting in $v_{\text{air}} = 20.1\sqrt{(273 + 18)} = 343$ m·s^{-1}.

The Equation 3.6 represents a non-linear function in which the speed of sound in air is proportional to the square root of the absolute temperature. Over the temperature range of interest, viz. 0°C and 40°C, we can approximate this relation to a linear equation,

$$v_{\text{air}} \approx 331.3 + 0.606T \tag{3.7}$$

where T is the temperature of the air in °C and v_{air} is given in m·s^{-1}. With this linearization we have a rate variation of approximately 0.6 m·s^{-1} for each 1°C increase.

3.1.1.2 Attenuation due to propagation

Sound propagation in the air is mainly attenuated by two factors: spherical spreading (also known as dispersion) and atmospheric absorption [65].

Spherical Spreading

Spherical spreading is the attenuation due to dispersion of sound waves when traveling in space and it has only relation with the propagation geometry. Sound waves may be seen as a disturbance that propagates in three dimensions. The simplest sound source model we can consider is a pulsating sphere, i.e. a sphere whose radius varies sinusoidally with time. If we consider a medium that is infinite, homogeneous and isotropic, the sound source model introduced before will produce a spherical wave that spreads out as it travels away from the radiating source [66], cf. Equation 3.8 . This occurs because the energy is being spread over a increasing spheric surface area making the wave front of the sound wave weaker.

$$A_{\text{sphere}} = 4\pi r^2 \tag{3.8}$$

The sound intensity is defined as the power per unit area and is given by,

$$I = \frac{W_{\text{source}}}{A_{\text{sphere}}} = \frac{W_{\text{source}}}{4\pi r^2} = \frac{k}{r^2} \tag{3.9}$$

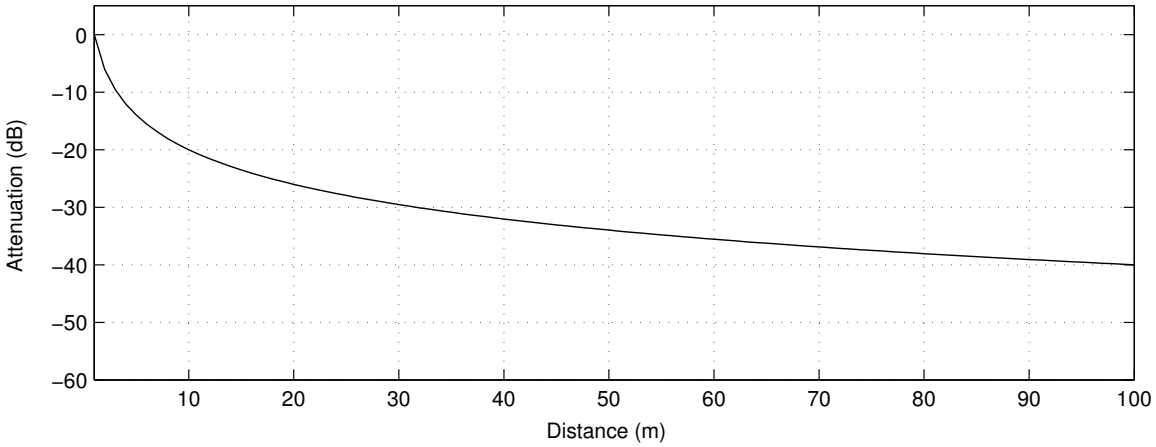
where I is the sound intensity (in $\text{W}\cdot\text{m}^{-2}$), W_{source} is the power of the source (in W), r is the distance to the source (in m) and k is a constant term given by $k = W_{\text{source}}/4\pi$. From Equation 3.9 we can observe that the sound intensity from a pulsating sphere in an infinite, homogeneous and isotropic medium reduces with the square of the increasing distance. This observation is known as as the inverse square law and is illustrated in Figure 3.2a.

Atmospheric Absortion

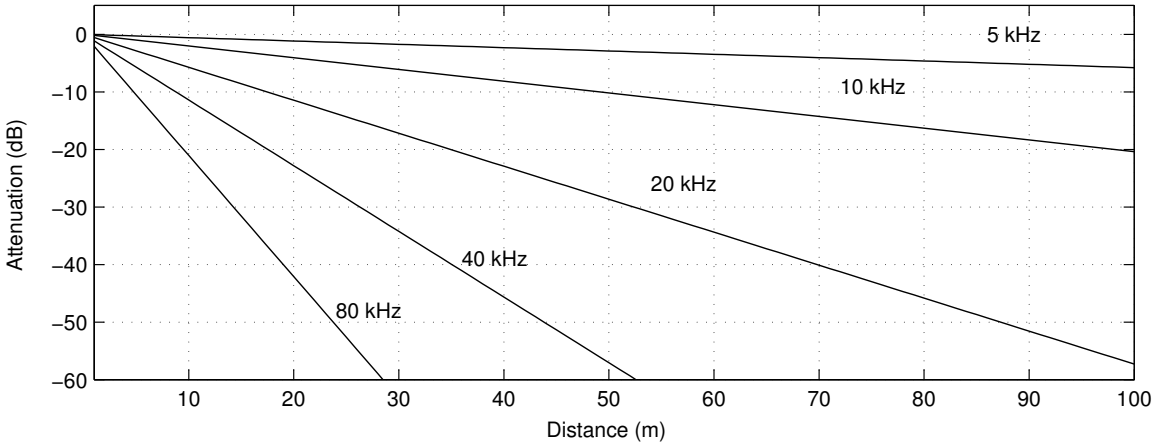
There are several sources of acoustic absorption in the air, such as dust and water molecules, that have greater impact at high frequencies and, as a result, sound not only gets attenuated but also gets duller as one moves away from a source. The amount of attenuation is dependent on the level of impurities and humidity, and is therefore variable [65].

In [67] and [68], Bass et al. introduced a simplified formula with corrections and improvements to its previous version [69] to obtain the atmospheric attenuation of sound due to the vibrational relaxation times of oxygen and nitrogen moleculas. This atmospheric absorption coefficient can be written as a function of frequency, relative humidity and atmospheric pressure, cf. Equation 3 from reference [68].

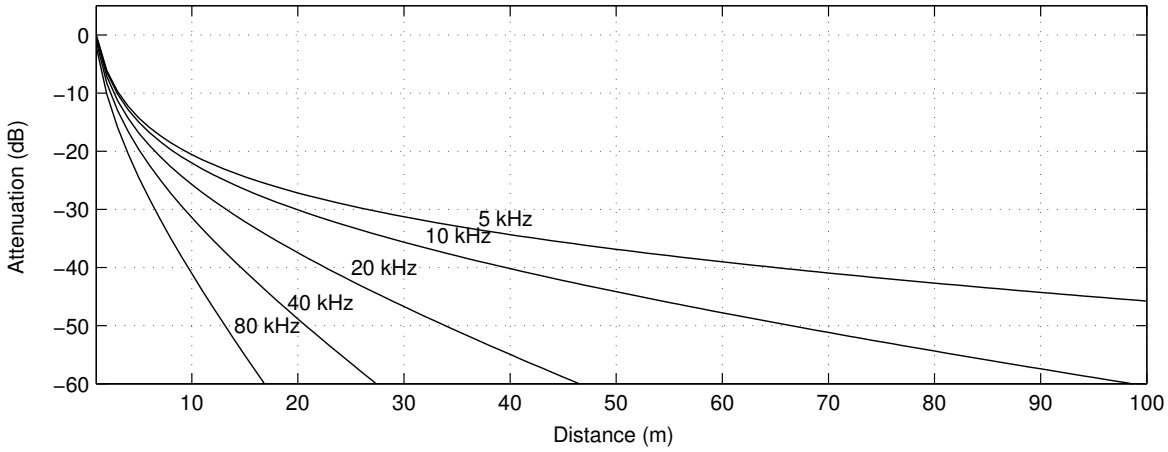
Based on the atmospheric absorption formula presented in [67] and [68] and considering T = 18°C, a relative humidity of H = 40% and an atmospheric pressure P_{atm} = 1 atm, a frequency dependent attenuation of the air with distance is presented in Figure 3.2b.



(a) Acoustic attenuation in dB due to spherical spreading over distance



(b) Acoustic attenuation in dB due to atmospheric absortion for different frequencies and considering $T = 18°C$, a relative humidity of $H = 40\%$ and an atmospheric pressure of $P_{atm} = 1$ atm.

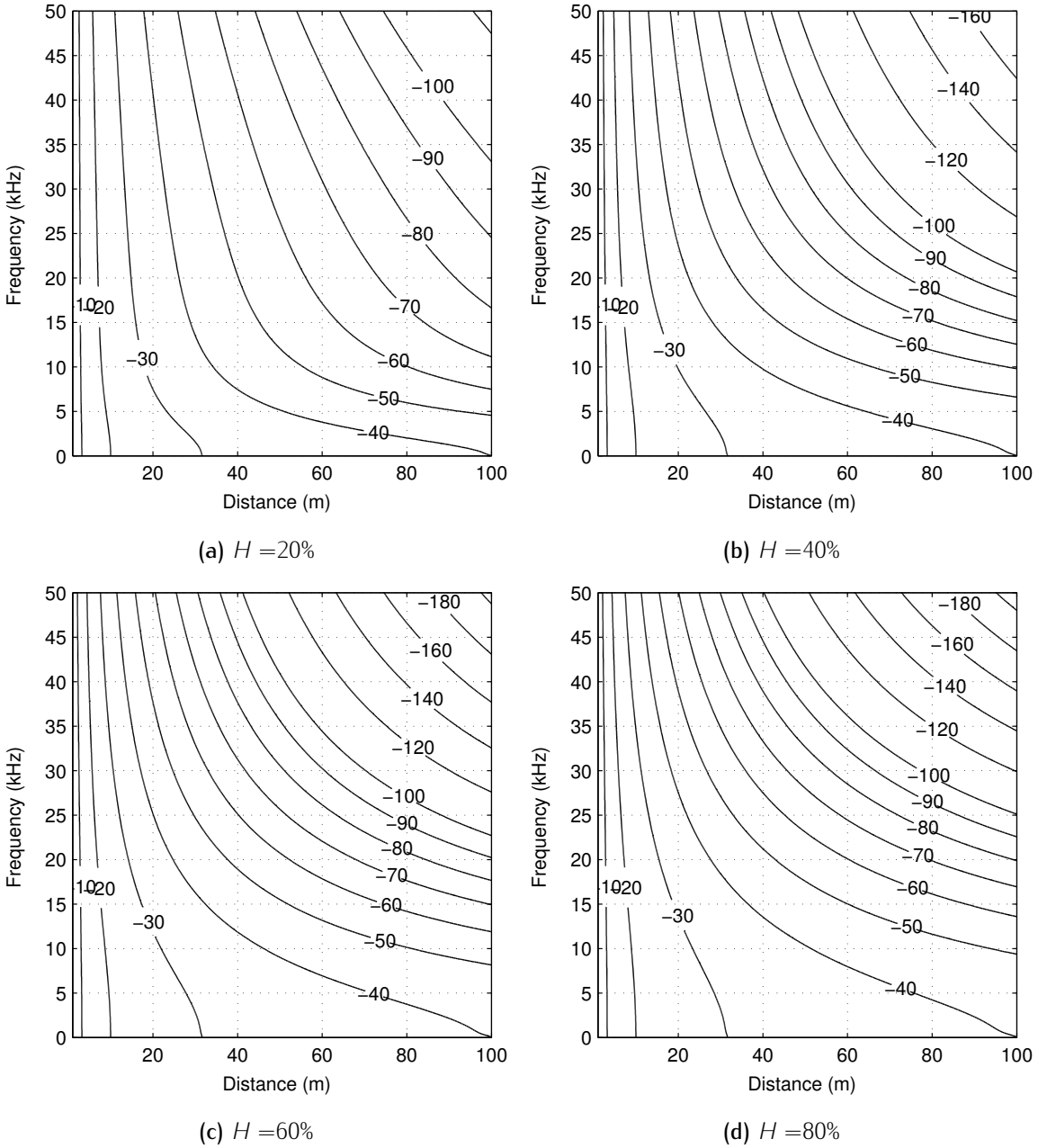


(c) Total attenuation in dB when we consider spherical spreading and atmospheric absortion.

Figure 3.2: Acoustic attenuation due to spherical spreading and atmospheric absorption.

Total Attenuation

The total attenuation of sound during propagation in the air can be obtained by the contribution of both factors introduced before (Figure 3.2c, that was obtained by summing the plots (a) and (b) from Figure 3.2). Figure 3.3 illustrates the total attenuation in dB for frequencies up to 50 kHz and ranging distances up to 100 m, considering a temperature of T = 18°C, an atmospheric pressure P_{atm} = 1 atm for different relative humidity values.



(a) H =20%

(b) H =40%

(c) H =60%

(d) H =80%

Figure 3.3: Total acoustic attenuation in the air due to atmospheric absorption and spherical dispersion considering T = 18°C and an atmospheric pressure of P_{atm} = 1 atm for distinct relative humidity cases.

3.1.2 Transducer Model

In section 3.1.1.2 we considered, as the sound source, a pulsating sphere transmitting homogeneously in the three dimensions. This raw and simple model is far from being a good approximation to what is really happening with a real transducer. A real transducer has a specific radiation pattern that heavily influences the transmitted or received acoustic energy. The radiation emitted by a transducer is different for the near and far field cases. Naturally, we are interested in the far field case, i.e. in the planar wave region.

In the reference theoretical model we follow, the transducer radiates like a planar circular piston in a three dimensional plane [66], cf. Figure 3.4. For the circular piston we will only focus on the 2D model because it can be easily extrapolated to the 3D model.

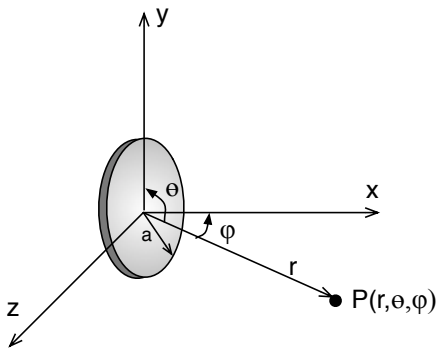


Figure 3.4: Transducer model using a circular plane piston geometry for the far-field case.

The beam function for this type of transducer can be approximated by the normalised beam pattern function given by the equation

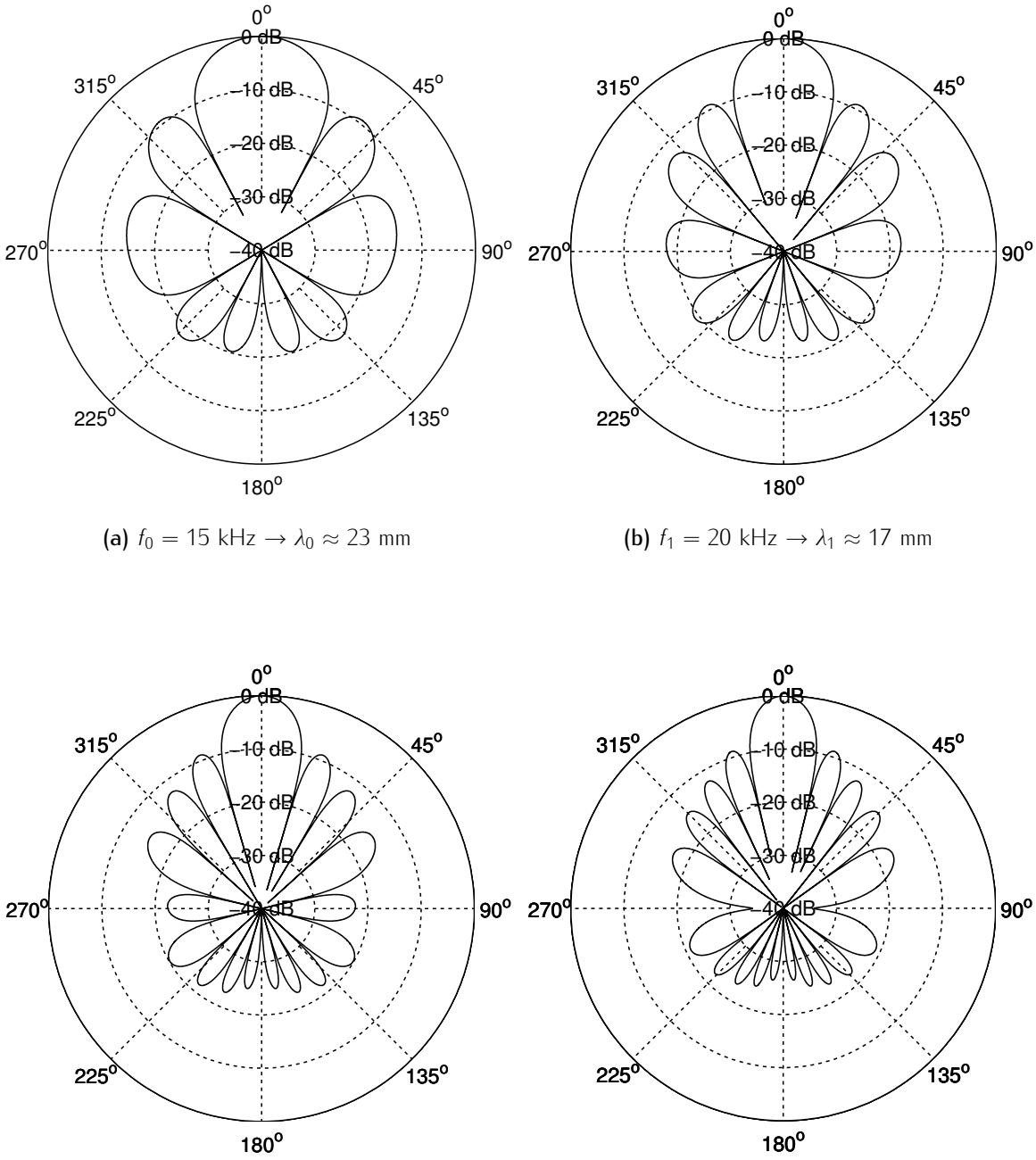
$$H_{3D}(a, \lambda, \theta, \varphi) = H_{2D}(a, \lambda, \theta) = 2\frac{J_1(\frac{2\pi a}{\lambda} sin\theta)}{\frac{2\pi a}{\lambda} sin\theta} \tag{3.10}$$

where J_1 is the first-order Bessel function of the first kind and λ is the wavelength of the radiated wave.

The previous model is for a circular piston pulsating in the air. Real world transducers have a hard case protecting the piston, which heavily attenuates the signal radiation on the sides and on the back of the transducer. To model this behaviour a raised-cosine function is used [66] which results in increased directionality of the transducer beam pattern.

$$H_b(a, \lambda, \theta, b) = H(a, \lambda, \theta)(b + (1 - b)cos\theta) \tag{3.11}$$

The beam pattern of the proposed transducer model (Equation 3.10), is plotted in Figure 3.5, for distinct frequencies.



(a) $f_0 = 15$ kHz $\rightarrow \lambda_0 \approx 23$ mm

(b) $f_1 = 20$ kHz $\rightarrow \lambda_1 \approx 17$ mm

(c) $f_2 = 25$ kHz $\rightarrow \lambda_2 \approx 14$ mm

(d) $f_3 = 30$ kHz $\rightarrow \lambda_3 \approx 11$ mm

Figure 3.5: Transducer model beam pattern using Equation 3.11 for different frequencies/wavelengths when $v = 343$ m·s^{-1}, i.e. $H_b(30\text{ mm}, \lambda_k, \theta, 0.5)$.

3.1.3 Noise Characterization

Acoustic background noise is present both in the audio and ultrasonic frequency ranges. Indoor acoustic background noise changes over time due to external disturbances of the acoustic environment. For example, people moving and talking inside a room or electronic devices operating produce broadband acoustic noise in the audible and ultrasonic bands. In acoustic ranging, due to the noise variability over time, it is important to have a real-time noise level estimate to adapt the ranging system and increase the system performance, viz. distance range and update rate improvements. Figure 3.6 presents the overall time domain ranging model with noise added.

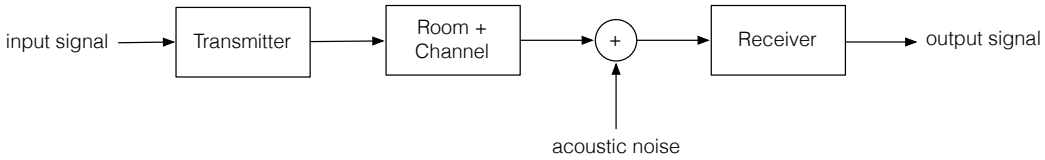


Figure 3.6: Ranging model with noise added.

Background noise can be defined as the total noise from all sources other than a particular one of interest [70]. Background noise is always present and its frequency content and power can change over time. Noise can be classified in the following three major types [71]:

Continuous: noise produced without interruption in a continuous mode, Examples are: machinery, such as blowers and pumps. Measuring for just a few minutes is sufficient to obtain the overall noise level and its correspondent frequency spectrum.

Intermittent: noise produced with interruptions and in cycles (events). This type of noise can be measured for each cycle just as for continuous noise. The cycle duration parameter must be taken into account. To measure the noise of an event, one must combine the level and duration into a single descriptor and the maximum sound pressure level may be used. The use of measures of a number of similar events can be used to obtain a noise average result.

Impulsive: noise produced with impacts or explosions that are brief and abrupt. Examples are: pile driver, punch press or gun-shot. The impulsiveness of noise may be quantified by looking to the difference between a quickly responding and a slowly responding parameter. Another important parameter that must be measured is the repetition rate (number of impulses per second, minute, hour or day).

If we consider a controlled acoustic environment, free of impulsive and intermittent noise types, the background noise level tends to remain constant (continuous noise type) and can be modeled with Additive White Gaussian Noise (AWGN). In this case, a propagating sound wave attenuates over distance due to spreading and absorption and the receiver SNR will decrease with the increasing distance that the wave travels. This factor limits the acting range of the system and results in a minimum acceptable SNR at the receiver. In this case, noise imposes a lower SNR limit that must be taken into account when designing the ranging system.

A common noisy type indoors is impulsive noise. This type of noise can be caused by mechanical environmental perturbations such as clapping hands, shaking a bunch of keys, collisions and friction between objects, etc. This impulsive events normally have higher power than the background noise, occur sporadically and are characterized by its small duration.

3.1.4 Room Model

We shall study a simple box-shaped room model, i.e. four walls plus floor and ceiling. When boundaries are considered, multipath sound propagation occurs, which negatively impacts the performance of an acoustic ranging system. We model an enclosed space with an internal volume $V = L_x \times L_y \times L_z$ (measured in m^3) and a total boundary surface $S = \sum_i S_i$ (measured in m^2). Each surface area can be assigned a distinct absorption coefficient ρ_a. In this model the reflected energy is a fraction of the incident wave, and is computed from the reflection coefficient, $\rho_r = 1 - \rho_a$.

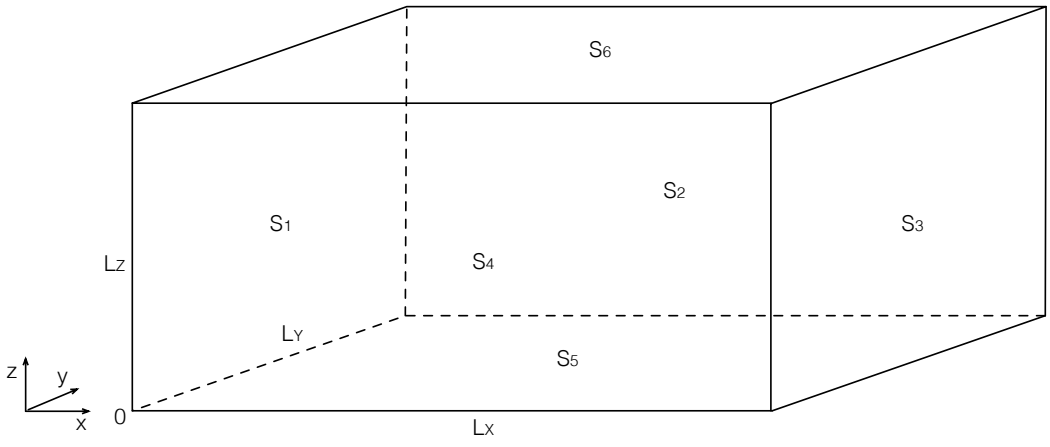


Figure 3.7: Enclosed space in the form of a box-shaped room.

Time-based ranging methods are heavily affected by multipath propagation. In indoor environments, NLoS propagation is a well known problem that must be modelled and mitigated, to increase the performance of the measurement stage. The standard discrete-time model for decomposition of a received multipath signal $y[n]$ is given by,

$$y[n] = \sum_i A_i s[n - \tau_i] + w[n] \tag{3.12}$$

where A_i is the complex amplitude of the ith path, τ_i is the delay of the ith path, $s[n]$ is the transmitted signal and $w[n]$ is AWGN. From above we can define the ith path as an isolated signal of the form $p_i[n] = A_i s[n - \tau_i]$.

Figure 3.8 illustrates the direct sound and 1^{st} order reflections when a transmitter and a receiver are inside the box-shaped room. All of these early reflections are perceived by the listener as part of a single sound.

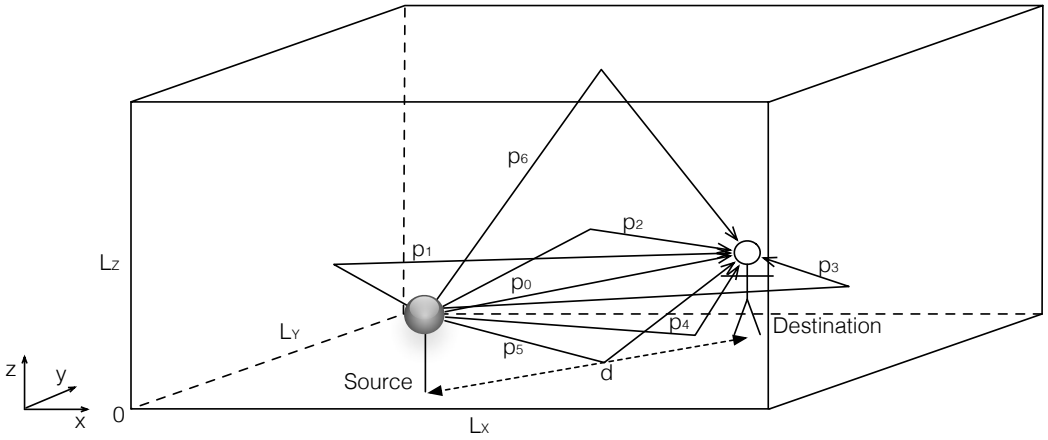


Figure 3.8: Direct sound and 1^{st} order reflections in a box-shaped room.

Figure 3.9 illustrates the multipath propagation of acoustic signals in indoor environments. The first pulse that arrives at the receiver is known as direct sound (p_0), being followed by the early reflections (Figure 3.8) and the diffuse reverberation tail that is commonly characterized by a sharp envelope.

In ranging applications, ideally we need to find the instant of arrival of the direct sound, which can be difficult due to the NLoS situations that may occur. In this case it is important to include in the model the early reflections stage, being the diffuse reverberation tail dispensable due to its noisy nature. A common way to reduce the diffuse reverberation tail impact is to include a guard time between measurements.

Figure 3.9 also illustrates the RT_{60} figure of merit that is used to measure the reverberation in large rooms. RT_{60} can be seen as the time required for reflections of a direct sound to decay 60 dB. An empirical formula, known as Sabine-Franklin's formula [66] defines RT_{60} as:

$$RT_{60} = \frac{24 \ \ln(10) \ V}{v_{air} \sum_{i=1}^{6} S_i(1 - \beta_i^2)} \tag{3.13}$$

where v_{air} is the speed of sound, V is the volume of the room, and β_i and S_i denote the reflection coefficient and the surface of the ith wall, respectively.

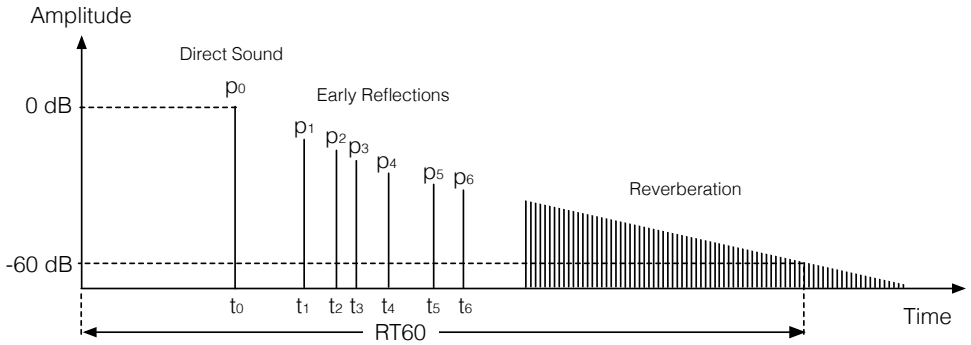


Figure 3.9: Time-domain representation of the room acoustics.

3.1.5 Room Acoustics for Simulation

Acoustic models can be used to predict the acoustic Room Impulse Response (RIR) for a given room setup. If the acoustic transmitter positions are known and additional information about the room is available (i.e. size, shape and wall surface materials) it is possible to obtain a RIR by using a specific microphone model placed at a specific position in the room. The RIR can be found by using an omnidirectional receiver model placed at the position of interest.

For simulation purposes we decided to use the indoor acoustics simulator proposed by Albuquerque et al., in [72]. The simulator is available online[1] and was developed with ultrasonic applications in mind. It uses a geometrical approach that considers all the wave reflections as specular which is a good approximation to the way the sound waves propagate in a closed space, i.e. the dimensions of the room are large compared to acoustic wavelengths [72]. For the frequency range that we are interested, i.e. frequencies above 10 kHz, the wavelengths are $\lambda < 3.6$ cm, so the smallest room dimension L for typical rooms is several orders of magnitude larger than λ, i.e. $L \gg \lambda$.

The simulator proposed in [72] by Albuquerque et al. uses a hybrid approach that combines the best features of the Ray Tracing Method and the Image Source Method [73]. In this case the Image Source Method is used to find all existent virtual sources followed by its validation using the Ray Tracing Method based only on the rays that are created by the virtual sources. If there is an obstructing surface on the path of the ray, the virtual source is invalid. This is called the visibility test and can be done by tracing the ray back [72].

For a certain configuration of the transducers, the simulator evaluates the acoustic RIR. This way, it is possible to test the influence of the room acoustics into an acoustic ranging system. The simulator has a modular structure, which simplifies the physical model expansion to meet our application criteria. The core code is implemented in Matlab and allows the inclusion of external

[1] `http://code.google.com/p/ultrasonic-simulator/`

models, such as specific transducer models. The selection of this simulator was influenced mainly by its simple but powerful features which include: wall reflections, sound attenuation and the transducer beam pattern [72].

3.1.5.1 Room Modeling

The room used for experimental validation in this work is Lab 4.3.17 in the Department of Electronics, Telecommunications and Informatics of the University of Aveiro. The room has dimensions $9 \times 8 \times 3$ m, concrete walls and ceiling, with large windows, linoleum floor and different types of office furniture (desks, chairs, closets, desktop computers and luminaires). Figure 3.10 presents a detailed floor plan of the room with transmitters and receivers placed for two example experiments, both with virtual walls added inside.

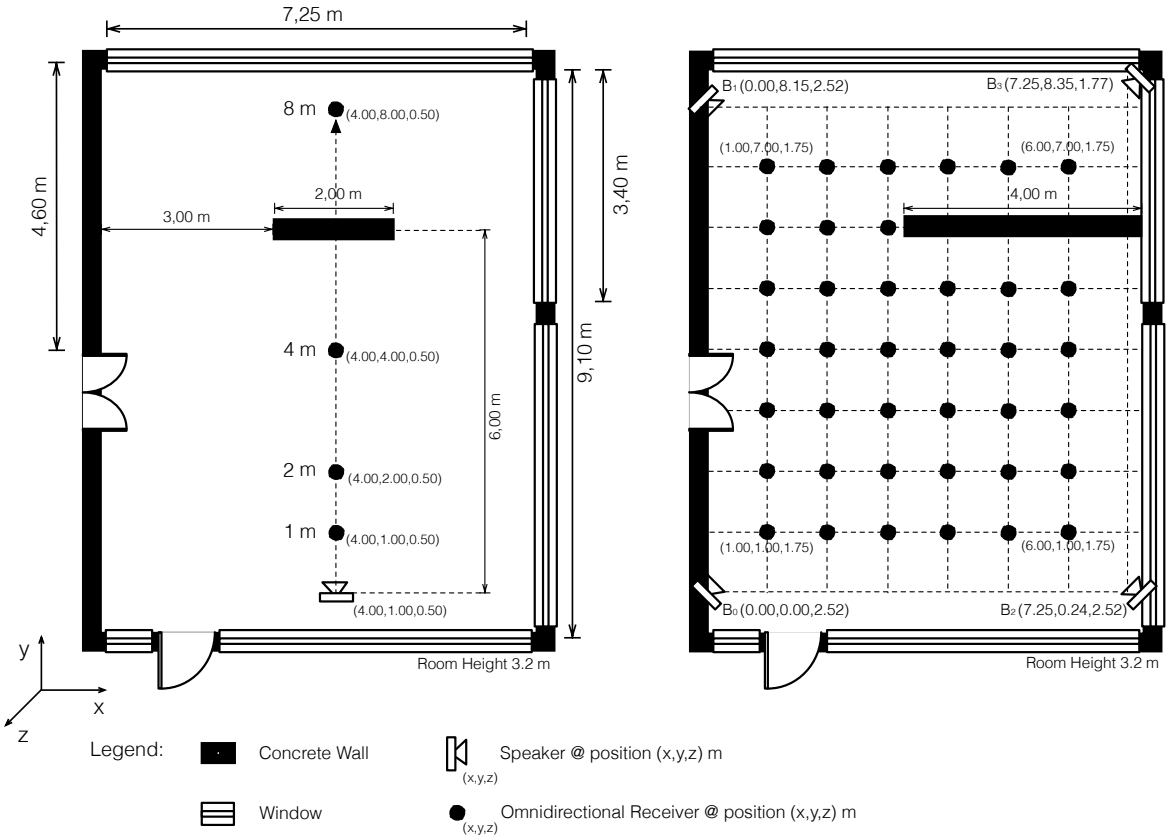
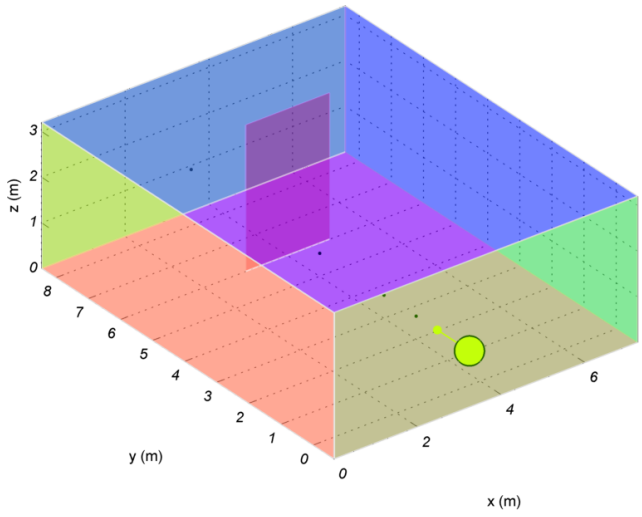
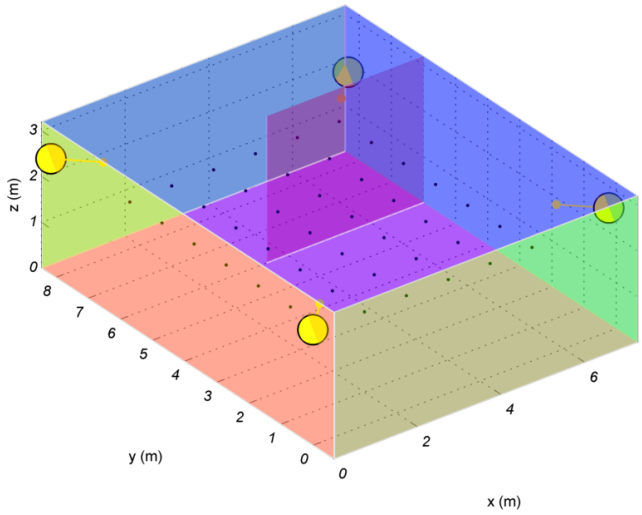


Figure 3.10: Lab 4.3.17 room floor plan with two experimental setups.

Using the simulator proposed in [72] and the models previously introduced, a room model for the two previous experiments is illustrated in Figure 3.11. The black dots are the positions where the receiver will be placed for RIR acquisition and the coloured circles represent the transmitters.



(a)



(b)

Figure 3.11: Box-shaped room models for the experiments presented in Figure 3.10.

3.1.5.2 Room Impulse Response

For each of the box-shaped room models presented in Figure 3.10 a set of Room Impulse Responses (RIRs) was obtained for the positions assigned. Based on this room model a series of RIRs was obtained using an omnidirectional receiver model moving in a grid of 6 m $\times$ 7 m, cf. Figure 3.10. For each position in the grid a set of four RIRs (one for each transmitting beacon) was obtained. These RIRs have a great importance for simulation purposes. Figure 3.12 shows the RIRs obtained for the example experiment defined in Figure 3.11a left.

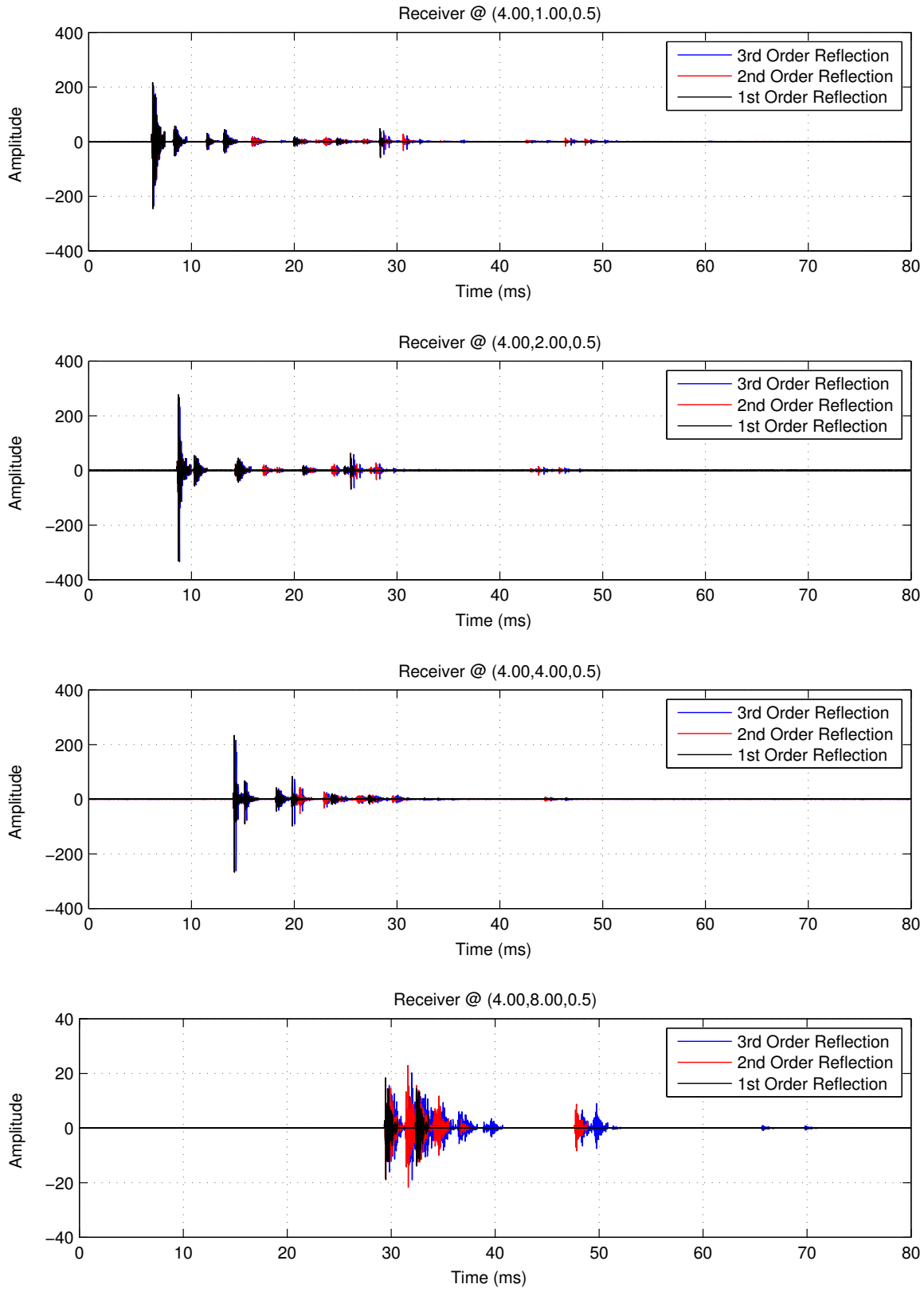


Figure 3.12: Noise-free RIRs for the room model presented in Figure 3.11a when the three first orders of reflection are considered, for distinct positions in the room. Temperature and relative humidity T=18° and H=40% were used in the simulation.

3.2 Perceptual Impact

In this section, a study on the perceptual impact of sounds on humans is undertaken, starting with the characterization of the human hearing system, frequency response and loudness sensation. Then, two non-invasive pulse design approaches are proposed: Time Domain Weighting and Perceptual Filtering. In the end, we describe the procedure used in the evaluation with subjects.

3.2.1 The Human Hearing System

The human hearing system allows us to hear and perceive the sounds of the environment around us. Additionally, it is also important in the regulation of our equilibrium. The ear receives sound waves and transforms them into electrical impulses, that are sent to the brain to be decoded and finally perceived. The hearing system can be divided into three parts, according to their function and their location. They are: the outer ear, the middle ear and the inner ear, cf. Figure 3.13.

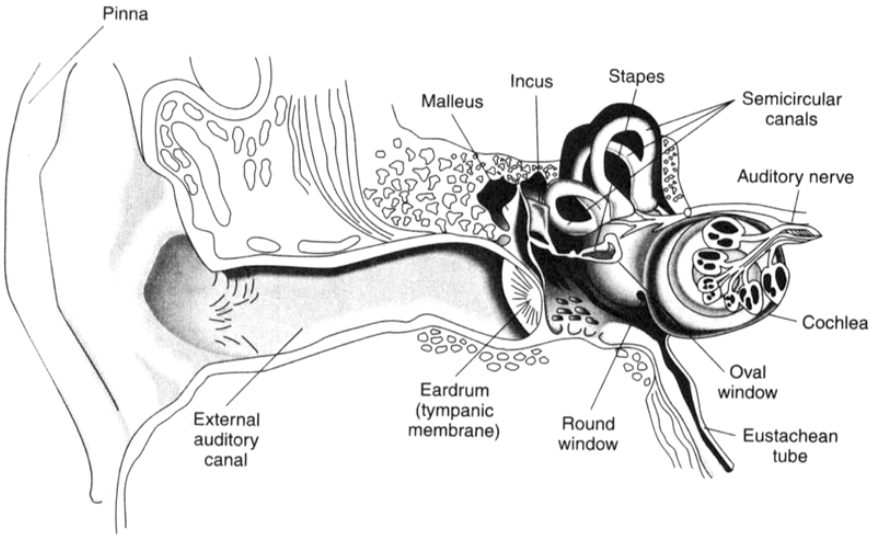


Figure 3.13: Human hearing system. Image from [74]

The outer ear includes the pinna and the external auditory canal, whose main roles are to collect, amplify and conduct sound waves to the eardrum. The middle ear is an air cavity behind the eardrum membrane, whereby energy of sound waves is transmitted to the cochlea through a structure composed by tiny bones. The inner ear is the most important part of the peripheral ear, being composed mainly by the cochlea, which is largely responsible for our ability to differentiate and interpret sounds. In fact, the cochlea performs a complex function by converting mechanical sound waves into electrical impulses that are forwarded to the brain via the auditory nerve.

The outer ear has an acoustic effect on sounds entering the ear canal, helping us to locate sound sources by modifying the frequency response of incoming sounds due to resonance effects, initially in the pinna and then in the auditory canal which main resonance frequency is around 4kHz. The tympanic membrane is a light, thin, highly elastic structure which forms the boundary between the outer and middle ears and is responsible to convert acoustic pressure variations into mechanical vibrations in the middle ear [65].

The two main functions of the middle ear are: (1) transmit the movements of the tympanic membrane to the fluid which fills the cochlea without significant loss of energy, and (2) to protect the hearing system to some extent from the effects of loud sounds, whether from external sources or the individual concerned.

The inner ear consists of the snail-like structure known as the cochlea. The function of the cochlea is to convert mechanical vibrations into electrical stimulus that are then transmitted to the brain. The cochlea consists of a tube coiled into a spiral which includes the basilar membrane that is responsible for carrying out the frequency analysis of incoming sounds, i.e. the cochlea performs a Fourier analysis of complex sounds into their component frequencies, acting as if it were made up of a bank of overlapping bandpass filters having bandwidths equal to the critical bands of the basilar membrane. A critical band can be defined roughly as the difference between two frequencies in order for a difference in pitch to be perceived.

3.2.1.1 Human Hearing Characterization

The human hearing system is commonly characterized based on its sensitivity and frequency range. It recognizes variations in pressure up to 6 orders of magnitude, from 10 μPa to 64 Pa and variations in frequency up to 3 orders of magnitude from 20 Hz to 20 kHz. Moreover, the existence of two ears enables 3D sound localization with high accuracy.

The classification of the audio band introduced in Figure 1.2 is directly related with the average human hearing frequency range of 20 Hz to 20 kHz. However, this frequency range can be quite different between individuals. It changes with the individuals age, particularly in the upper frequency limit where it tends to reduce. Healthy young children can have a full hearing frequency range up to 20 kHz, but at the age of 20, the upper frequency limit is likely to have fall to 16 kHz. This phenomenon is known in the literature as "presbycusis" [65].

Figure 3.14 depicts the average curves that show the decline in hearing sensitivity for both men and woman over age. These hearing losses increase for higher frequencies and have greater impact in men rather than in women.

The hearing sensitivity varies over a vast sound pressure level range for different frequencies. On average, the lowest sound pressure level that can be detected by the human hearing system is approximately 10 μPa around 4 kHz and the highest sound pressure level just below the threshold of pain is 64 Pa around 100 Hz. The ratio between these thresholds has several orders of magnitude and can be obtained by:

$$\frac{\text{Threshold of pain}}{\text{Threshold of hearing}} = \frac{T_P}{T_H} = \frac{64}{10^{-5}} = 6400000 = 6.4 \times 10^6 \tag{3.14}$$

From Equation 3.14 we can observe a very wide range variation of pressure levels. To deal with values of such magnitude the sound pressure level (SPL) is represented in decibels with relation to $P_{ref} = 20\ \mu$Pa (average human hearing lowest sensitivity around 1 kHz) as dB (SPL) as follows:

$$\text{SPL} = 20 \log_{10} \left(\frac{P}{P_{ref}} \right) \ dB\ (SPL) \tag{3.15}$$

where P is the pressure level and P_{ref} is the pressure reference level (20 μPa).

Therefore, we can obtain the threshold of hearing and the threshold of pain in dB (SPL) from

$$T_H = 20 \log_{10} \left(\frac{2 \times 10^{-5}}{2 \times 10^{-5}} \right) = 0 \text{ dB (SPL)} \tag{3.16}$$

$$T_P = 20 \log_{10} \left(\frac{64}{2 \times 10^{-5}} \right) = 130 \text{ dB (SPL)} \tag{3.17}$$

This approach gives us a more appropriate range scale of values, from 0 to 130 dB (SPL). Magnitudes between 0 and 100 dB (SPL) are commonly observed in everyday situations.

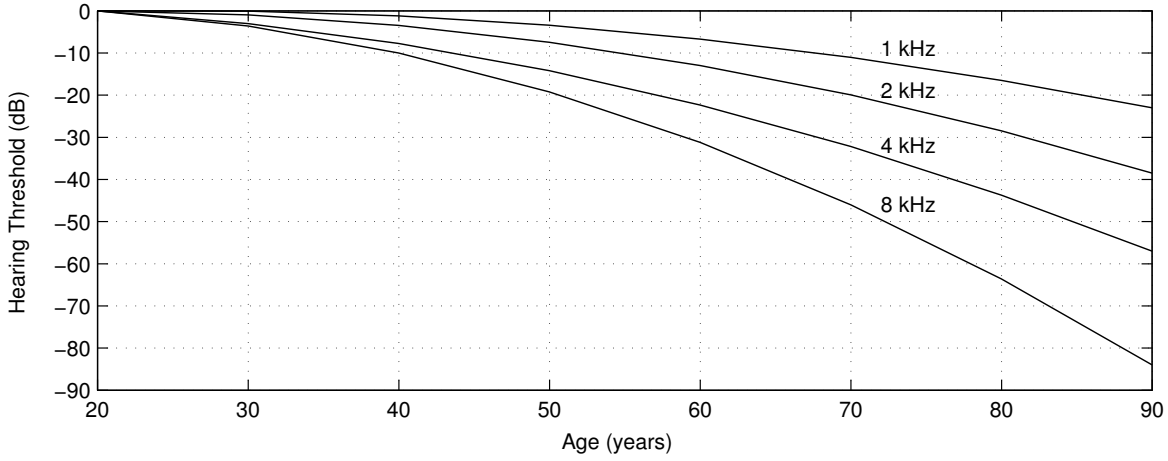


Figure 3.14: Human hearing sensitivity average loss with age. Data from [65]

3.2.1.2 Safety Limitations

Ultrasonic sound is beyond the range of normal human hearing and is usually defined as sound with frequencies above 20 kHz that can not be perceived by the humans. However, the human ear sensitivity shows a much wider range than the usually cited 20 Hz to 20 kHz frequency audio band. Although these sounds are imperceptible for humans they present an invisible health risk due to the possibility of damage of the hearing system. To prevent this damage, safety limits are used to reduce the exposure to these high frequency acoustic waves.

In [75] the authors make a detailed description of the effect of ultrasound on humans. They state that for frequencies above 17 kHz and at levels above 70 dB (SPL) adverse subjective effects, i.e. fatigue, headache and malaise, have been experienced. Moreover, they state that these subjective effects are highly correlated with the hearing capabilities of the subject, and report that younger individuals are more affected than older ones, and that the effects are more clearly experienced by women than men. Table 3.1 presents several sets of recommended ultrasound exposure limits for 1/3 octave bands in SPL, for the range 8–50 kHz [76].

	Proposed By							
	a)	b)	c)	d)	e)	f)	g)	
Freq. (kHz)	SPL (dB re 20μPA)							Lower Bound
8	90	75	-	-	-	-	-	75
10	90	75	-	-	80	-	-	75
12.5	90	75	75	-	80	-	-	75
16	90	75	85	-	80	-	75	75
20	110	75	110	105	105	75	75	75
25	110	110	110	110	110	110	110	110
31.5	110	110	110	115	115	110	110	110
40	110	110	110	115	115	110	110	110
50	110	-	110	115	115	110	110	110

Table 3.1: Recommended sound/ultrasound exposure limits in octave bands. Proposed by: a) Japan (1971); b) Acton (1975); c) USSR (1975); d) Sweden (1978); e) American Conference of Governmental Industrial Hygienists (ACGIH 89) and US Department of Defense (2004); f) International Radiation Protection Agency (IRPA 1984) and g) Canada (1991). Data from [76]

We compile in the rightmost column the lower bound among all the proposed sets of data. This resulted in a SPL value of 75 dB for frequencies in the range 8–20 kHz and in a 110 dB (SPL) for frequencies in the range 25–50 kHz. Taking a conservative point of view, we chose to limit the maximum transmitted SPL value to 75 dB in the band of interest, i.e. between 15–25 kHz.

3.2.2 Loudness Perception

Loudness is the subjective perception of the intensity of a sound by a specific listener. There is no objective measure of loudness and measurement procedures should be based on statistical methods to converge upon an evidence for indirectly measure loudness. Loudness is a subarea of psychoacoustics that studies the relationship between the physical properties of a sound and the way they are perceived by the listener. In a first approach loudness has a strong relation with the level of a sound but it also changes with other physical properties of the sound stimulus, such as: frequency, bandwidth, duration (temporal integration), spectral complexity, presence of other sounds (partial masking), and physical environment (noise). Furthermore, loudness can also be affected by other subjective factors that have nothing to do with the stimulus, e.g. psychological and physical state of the listener, changes with memory over time, subject criterion (particular cognitive factors), subject differences (old vs. young; male or female), the way stimulus is presented, the way loudness is measured, multisensory context, and cross-cultural differences [77].

The first steps toward measuring the loudness of sounds came in the second half of the nineteenth century when researchers were trying to quantitatively measure the perceived magnitude of a perceptual experience in relation to the magnitude of the physical stimulus that evokes the experience. In 1743 Johann Krüger had proposed for the first time the relation between the intensity of sensation and the intensity of a physical stimuli responsible for the sensation. A century later Fechner would recognize that there may be quantitative as well as qualitative differences between stimuli and sensations. He has proposed a psychophysical law in which he stated that the magnitude of a subjective sensation increases proportional to the logarithm of the stimulus intensity [77]. This observation was instrumental in the study of loudness measurement and has lead to the adoption of the decibel notation to represent relative values of sound intensity or sound pressure.

During the twentieth century several researchers have extended the work of Fechner by proposing methods and procedures for the quantification of loudness. The work of pioneers such as Fletcher and Munson (1933) [78], Stevens (1936) [79] and Robinson and Dadson (1956) [80] are still cited today by researchers due to its wide acceptance by the scientific community as can be seen in several recent works, cf. [10].

The scaling of loudness is an important topic due to its relation to the physical magnitude of sound. Based on the assumption that loudness is always subjective, several methods have been proposed during the last century, of which we highlight the *sone* and the *phon* scales. The proposal for a loudness scale was pioneered by Stevens in 1936. In [79] he was able to pull together several

sets of findings and used them to construct a scale for loudness measurement that he called the *sone* scale. One sone is defined as the loudness of a 1 kHz tone at 40 dB (SPL), when presented binaurally in free-field from the frontal direction. A 1 kHz tone with a level of 50 dB (SPL) is normally perceived as about twice as loud as the 40 dB (SPL) tone which results in a loudness of 2 sones. This simple rule can not be applied for sound levels below 40 dB (SPL). At low levels the loudness changes more rapidly with sound level as can be observed in the Equation proposed by Stevens [79],

$$L = kI^{0.3} \tag{3.18}$$

where k is a constant depending on the units used in the measurements. This equation states that the loudness of a sound is proportional to its intensity raised to the power of 0.3.

It the twenties Barkhausen introduced the *phon* loudness scale. By definition the *phon* scale is represented by curves in which the sound for different frequencies sounds just as loud as a frequency of 1 kHz. These curves connect points of equal loudness and are known in literature as equal-loudness contours that represent equal-loudness levels over the audible frequency range in which sounds are perceived as equally loud by the average young person without significant hearing impairment [81].

Figure 3.15 presents the average relationship between loudness in sones and loudness level in phons for the ISO523B and ANSI S34:2007 standards. From the figure it is possible to observe the linear relation between the two scales (note that the sones scale is also in logarithm) for loudness levels above 40 dB (SPL) as stated before. Below the 40 dB (SPL) it is possible to observe the behaviour described by Stevens in [79], cf. Equation 3.18.

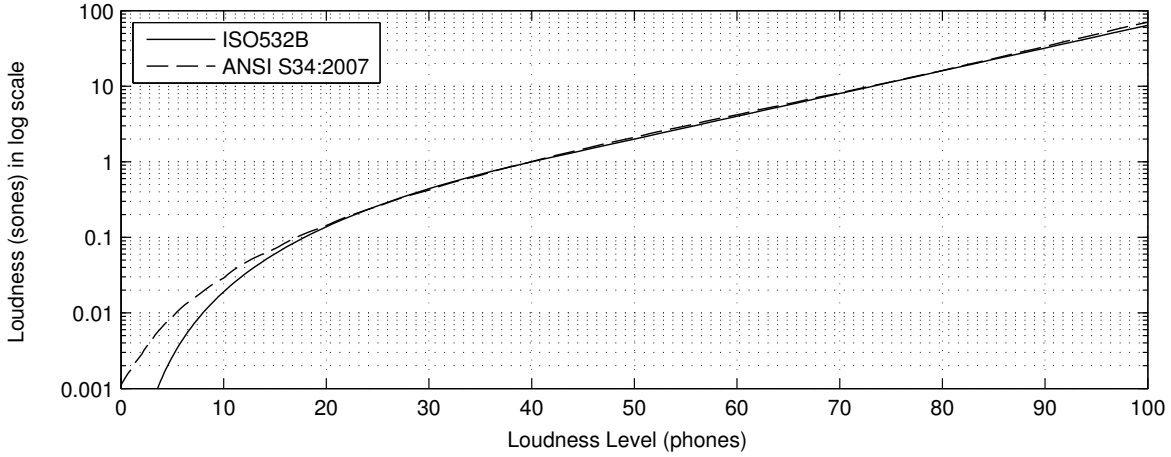


Figure 3.15: Loudness scales Sone Vs. Phon

3.2.2.1 Threshold of Hearing

In 1933 Fletcher and Munson suggested a novel approach for loudness measurement. They created a proper scale for loudness that was based in the principle of additivity by relating different amplitudes of the stimulus to a common tone at 1 kHz using headset measurements [78]. In 1956 Robinson and Dadson proposed a set of equal-loudness-level contours based in free-field measurements. The curves were determined experimentally and are reported in [80]. In 1987, the work of Robinson and Dadson was used as the basis for the standard ISO 226 definition, which specifies the equal-loudness contours for a listener in a free sound field. The specification for the threshold of hearing in ISO 226 is the basis for sound field audiometry and is described in the standard ISO 8253-2. The ISO 226 was revised in 2003 and as a result of discussion it was decided to divide the standard in two parts, one for free-field and the other for diffuse-field.

ISO 226:2003

The ISO 226:2003 revision specifies combinations of sound pressure levels and frequencies of pure continuous tones which are perceived as equally loud by human listeners [82]. The standard specification was based in the following measurement procedures: free field measurements; the sound source is directly in front of the listener; the sound signals are pure tones; the sound pressure level is measured at the position where the centre of the listener's head would be, but in the absence of the listener; listening is binaural and the listeners had normal hearing capabilities in the age range between 18 to 25 years old.

Figure 3.16 presents the standard ISO 226:2003 equal-loudness contours for *phon* values of 0, 20, 40, 60 and 80 dB. Note that the zero *phon* line represents the threshold of hearing contour. As the ISO 226:2003 standard is not defined for frequencies greater than 12.5 kHz, and given that we are designing non-invasive pulses with frequency content in the higher band of the audible range, an extended model with data up to 20 kHz is addressed in Section 3.2.3.

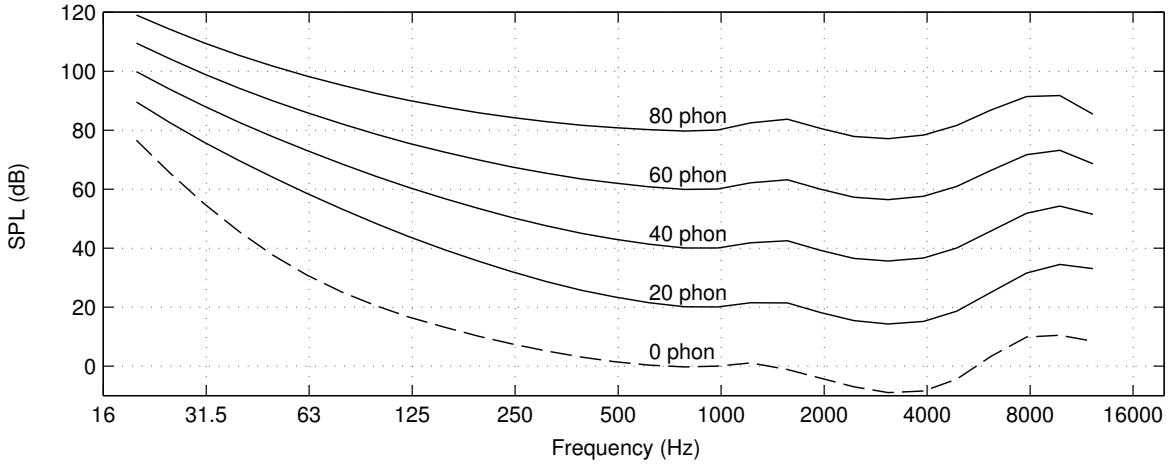


Figure 3.16: Equal loudness contours by *phon* (ISO 226:2003) [82].

3.2.2.2 Pulse Duration

Pulse duration has considerable impact on loudness perception. For tone durations above approximately 200 ms the perceived level does not change which allows us to infer that loudness contributions are integrated over time, with a time constant of the order of 100–200 ms [77], cf. Figure 3.17. However, when the tone pulse is shorter the perceived loudness reduces. In this case the perceived loudness is inversely proportional to the duration of the sound stimulus. This means that the sound amplitude variation cannot be significantly perceived for short durations, but by the sound level averaged over 200 ms. This variation is well known for frequencies in the range 1-4 kHz, but in general is more pronounced for greater frequencies [65]. This phenomenon is known in the literature as temporal integration.

In [83] the authors have studied the relation between the sound perception threshold and its duration. They concluded that the perception of sound showed an increase in loudness that is dependent with the increase in duration of the acoustic stimulus and in [84] the authors conduct several experiments to evaluate the binaural and temporal integration impact on loudness perception in the presence of pure tones and noises. In the experiments they used a set of tones and noise pulses, with varying intensity, duration and mode of presentation (monaural/binaural). They concluded that loudness increases with both increasing duration and increasing intensity and beyond the critical duration it becomes relatively independent of duration, which confirms the general nature of temporal integration.

When sound tones are reduced in duration to less than about 200 ms, their level must be raised to remain audible. The increment in sound level needed is about 10 dB for a reduction factor of 10 in sound duration. This dependence of threshold with the sound duration corresponds to a temporal integration of the sound intensity within a time window of 200 ms.

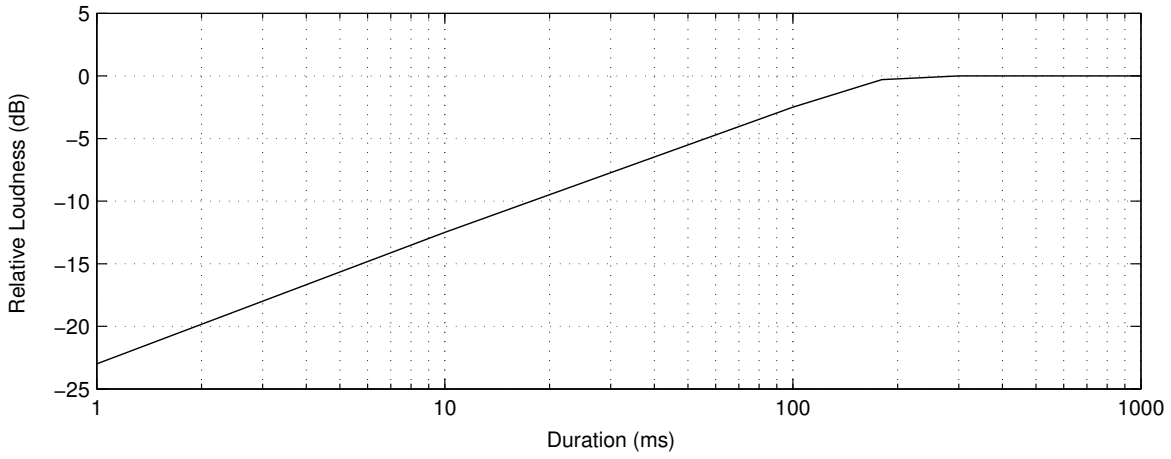


Figure 3.17: Effect of pulse duration on loudness perception (temporal integration) for frequencies in the range 1-4 kHz. Data from [65].

3.2.2.3 Pulse Bandwidth

The pulse bandwidth also affects the perceived loudness. For complex sounds with multiple frequencies it has been well documented that their spectral bandwidth plays an important role in loudness perception [77]. The loudness of a sound within a critical band is independent of the number of frequency components so long as their total acoustic intensity is constant. When the frequency components of the sound extend beyond a critical band, an additional effect occurs due to the presence of components in other critical bands. In this case more than one critical band is contributing to the perception of loudness and the brain appears to add the individual critical band responses together. This increases the perceived loudness even though the total acoustic intensity keeps unchanged [65]. Many studies have shown that the overall level of a narrow-band sound must be higher than the level of a broadband sound to be perceived as equally loud. This effect is typically referred to as spectral loudness summation [77].

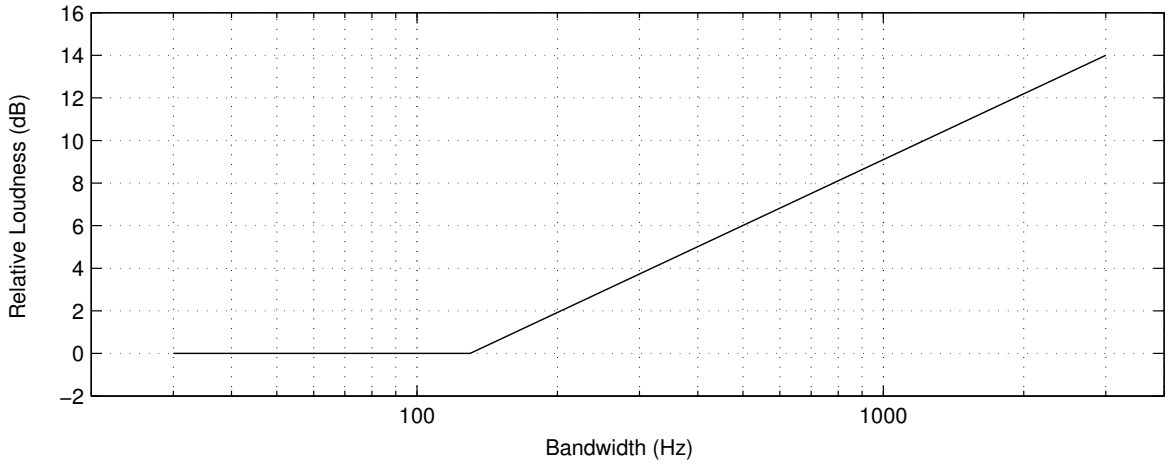


Figure 3.18: Effect of pulse bandwidth on loudness perception when a 1 kHz tone is judged equally loud to the level of a 60 dB complex five-tone around 1 kHz. Data from [77].

3.2.3 Perceptual Filter Design

In 2004 Suzuki and Takeshima presented in [10] a study which analysed the results of twelve contemporary studies to arrive to a new set of contours. The new set of contours they have estimated was compared with the two classic sets of data available in the literature. They concluded that the contours described by Fletcher and Munson exhibit some overall similarity to their estimated contours in the mid-frequency range up to 60 phons. Moreover, the contours described by Robinson and Dadson exhibit clear differences from the new contours mainly below 500 Hz, with a discrepancy of up to 14 dB. Figure 3.19 shows the results they obtained for the threshold of hearing (zero *phon*) when pure tones are used as stimulus. The authors estimate the threshold of hearing based on a smoothed line of the averages of the experimental data based on a cubic B-spline function for the

frequency range from 20 Hz to 18 kHz to obtain the threshold of hearing estimated contour [10].

The threshold of hearing contour estimated by Suzuki and Takeshima [10] will be the basis for our threshold of hearing model, which will be used to design an equivalent linear-phase filter. This filter will process the signals used in the non-invasive ranging process.

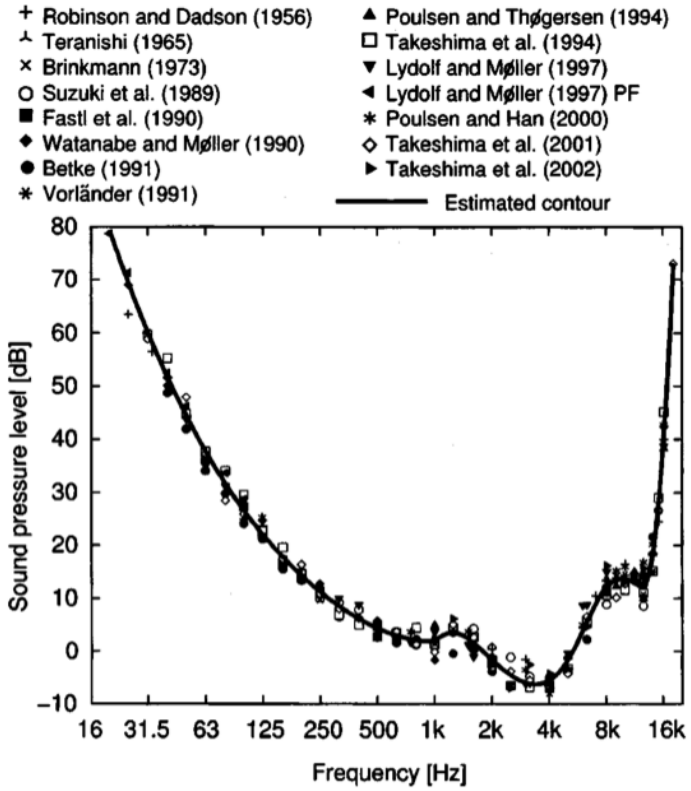


Figure 3.19: Equal loudness contour for the threshold of hearing estimated by Suzuki and Takeshima in 2004. Image from [10].

3.2.3.1 Zero-Phase Filter Approximation

A linear-phase filter is used when a causal filter is needed to modify the magnitude-spectrum of a signal while preserving its time-domain waveform as much as possible [85]. Linear-phase filters have a symmetric impulse response, such as

$$h(n) = h(-n), \quad n = -(N-1)/2, \ldots, -1, 0, 1, \ldots, (N-1)/2 \tag{3.19}$$

This symmetric-impulse-response constraint means that linear-phase filters must have an impulse response with finite duration, i.e. a FIR filter, because a causal recursive filter cannot have a symmetric impulse response. Every real symmetric impulse response corresponds to a real frequency response multiplied by a linear phase term $e^{-j\alpha\omega T}$, where $\alpha = (N-1)/2$ is the slope of

the linear phase. Linear phase is often ideal because a filter phase of the form $\Theta(\omega) = -\alpha\omega T$ corresponds to phase delay,

$$P(\omega) \equiv -\frac{\Theta(\omega)}{\omega} = -\frac{-\alpha\omega T}{\omega} = \alpha T = \frac{(N-1)T}{2} \tag{3.20}$$

and group delay

$$D(\omega) \equiv -\frac{\partial}{\partial\omega}\Theta(\omega) = -\frac{\partial}{\partial\omega}(-\alpha\omega T) = \alpha T = \frac{(N-1)T}{2}. \tag{3.21}$$

That is, both the phase and group delay of a linear-phase filter are equal to $(N-1)/2$ delay samples at every frequency. Since a length N FIR filter implements $N-1$ samples of delay, the value $(N-1)/2$ is exactly half the total filter delay. Delaying all frequency components by the same amount preserves the waveshape, in the time domain, as much as possible for a given amplitude response [85].

To model the threshold of hearing contour proposed by Suzuki and Takeshima (Figure 3.19) we have started by obtaining discrete points of the equal loudness contour as presented in Figure 3.20. Then a cubic interpolation of the available data with higher frequency resolution was performed and extrapolated from 18 kHz up to 20 kHz to obtain the positive part of the frequency response. In this case we only have the magnitude data (real values) of the positive frequencies that should be used to generate an even symmetry frequency response.

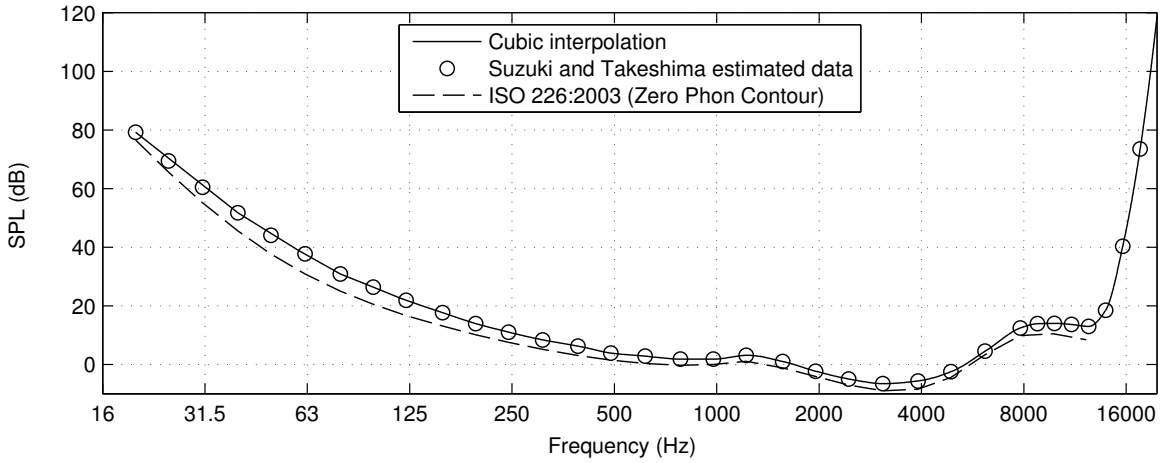


Figure 3.20: Extrapolation of the threshold of hearing contour estimated by Suzuki and Takeshima [10] for frequencies up to 20 kHz.

A way to obtain a FIR filter from the cubic interpolation of the threshold of hearing frequency response presented in Figure 3.20 is to design a zero-phase FIR filter. The real impulse response

$h(n)$ of a zero-phase filter is always even and satisfies the condition stated in Equation 3.19 which states that the impulse response must be symmetric about time zero.

A real frequency response has phase zero when it is positive, and a phase of π radians when it is negative. Therefore, a zero-phase filter, as we have previously defined in Equation 3.19, may actually have a phase response of 0 or π at each frequency. This type of filter has precisely zero-phase in all pass bands, while it switches between 0 and π in the stop bands, i.e. in the frequency bands where the gain is desired to be zero [85].

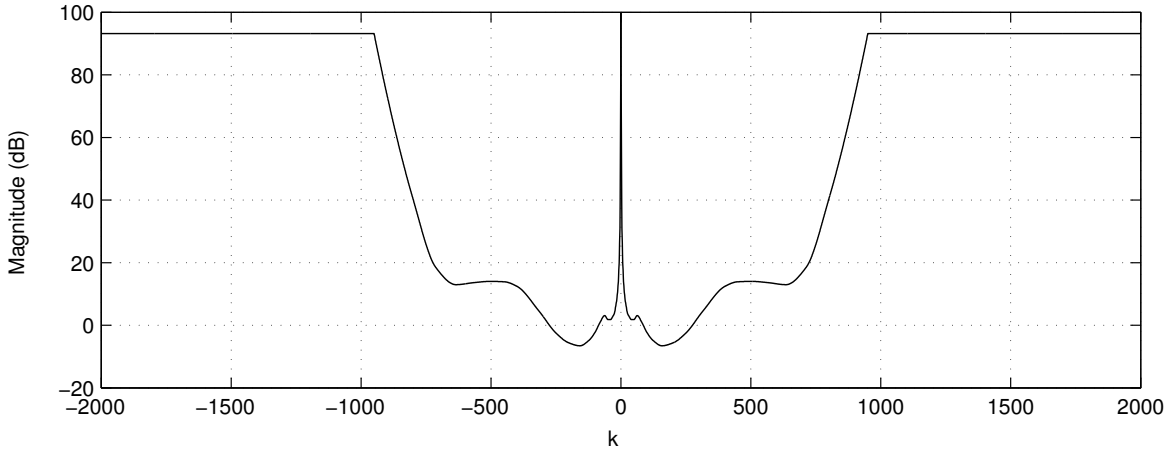
To obtain a real symmetric and non-causal impulse response $h(n)$ we have used the real part of the cubic interpolation frequency response with a constant phase of zero to generate $H(k)$. By forcing the imaginary part of $H(k)$ to zero we obtained a spectrum function with zero phase content. Assuming an even spectrum function with zero imaginary part, to generate $H(k)$ we need to guarantee spectral symmetry around the zero frequency, i.e. $H(k) = H(-k)$, as can be observed in Figure 3.21a. This ensures that a real symmetric and non-causal impulse response $h(n)$ is obtained.

This approach is ideal because we do not have any phase information in the original data. Then by performing the inverse of the discrete Fourier transform (IDFT), cf. Equation 3.22, of $H(k)$ we obtain a real and non-causal $h(n)$ with even symmetry, cf. Figure 3.21b.

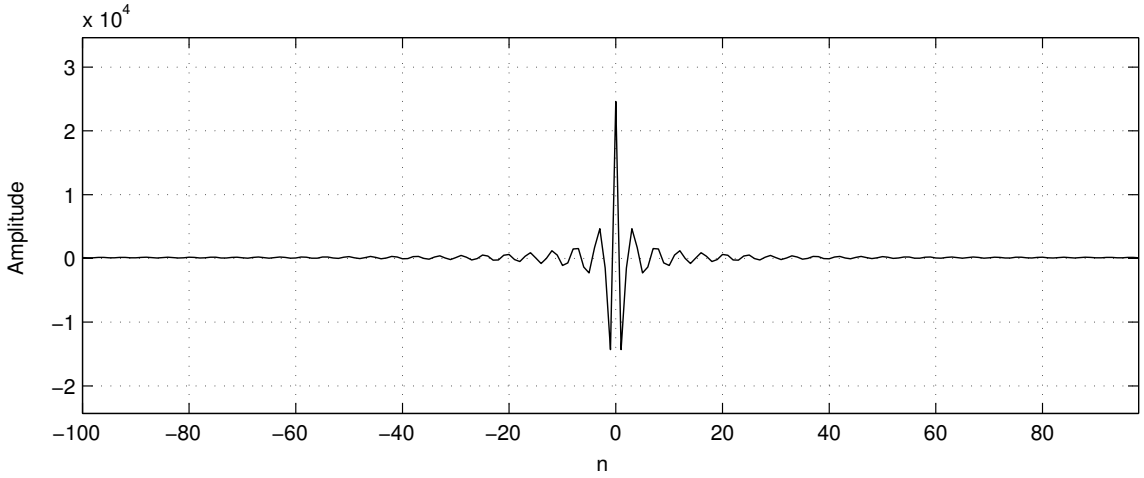
$$h[n] = \frac{1}{N}\sum_{k=0}^{N-1} H(k)e^{j2\pi nk/N} \quad \text{(IDFT)} \tag{3.22}$$

In Equation 3.22, $H(k)$ is a real and symmetric N-point frequency spectrum composed with interpolated data from the Suzuki and Takeshima estimated data, with Fourier frequencies k ranging from $-N/2+1$, through the zero frequency bin and up to the highest Fourier frequency bin $N/2$. Each bin number represents the integer number of sinusoidal periods present in the time series. The amplitudes and phases represent the amplitudes A_k and phases ϕ_k of those sinusoids. In summary, each bin can be described by $H(k) = A_k e^{j\phi k}$. The real part of the frequency response $H(k)$ is symmetrical around the zero frequency and the imaginary part of $H(k)$ is forced to zero thus resulting in a non-causal impulse response in the time domain, cf. Figure 3.21b.

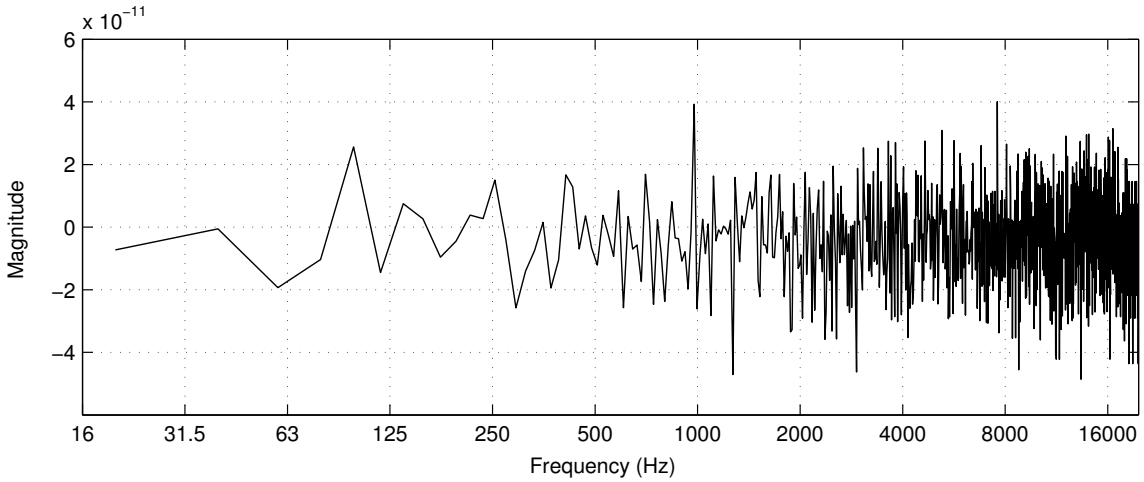
Figure 3.21b shows a truncated version (only for visualization purposes) of the impulse response obtained with Equation 3.22. The complete non-causal impulse response will be used as FIR filter coefficients to process the pulses used in non-invasive ranging. To validate the impulse response obtained ($h[n]$), we have applied to it the DFT, to go back to the frequency domain. Figure 3.21c shows the error obtained between the original Suzuki and Takeshima [10] frequency response and the frequency response of the zero-phase FIR filter approximation. As data will be processed off-line we have opted for maintaining a good resolution at low frequencies which resulted in a higher order impulse response, viz. N=4001, which is easily processed in a desktop computer.



(a)



(b)



(c)

Figure 3.21: Zero-phase FIR filter approximation of the threshold of hearing contour estimated by Suzuki and Takeshima [10]. (a) Real even symmetry frequency response $H(k)$ with N=4001 frequency bins. (b) Truncated even non-causal impulse response $h[n]$. (c) Error obtained between the original Suzuki and Takeshima [10] frequency response and the zero-phase FIR filter approximation.

3.3 Hardware Constraints

Hardware characterization of the receiver is important due to the constraints it will impose on the pulse design process. In particular, the available bandwidth and the processing capabilities must be taken into account.

3.3.1 Smartphone Compatibility

Several smartphones were tested to obtain their audio input bandwidth. The headset input in smartphones is prepared to be used with conventional low-cost electret microphones. This type of microphone need phantom power to work properly. To measure the available input bandwidth in the earplug of conventional smartphones a proper cable was constructed (Figure 3.22), to uncouple the signal generator and the headset audio input DC phantom power component. A first order passive HPF with cutoff frequency of approximately 34 Hz was used to decouple the devices, cf. Equation 3.23. Note that the smartphone headset input has a fourth contact normally used as an external microphone input.

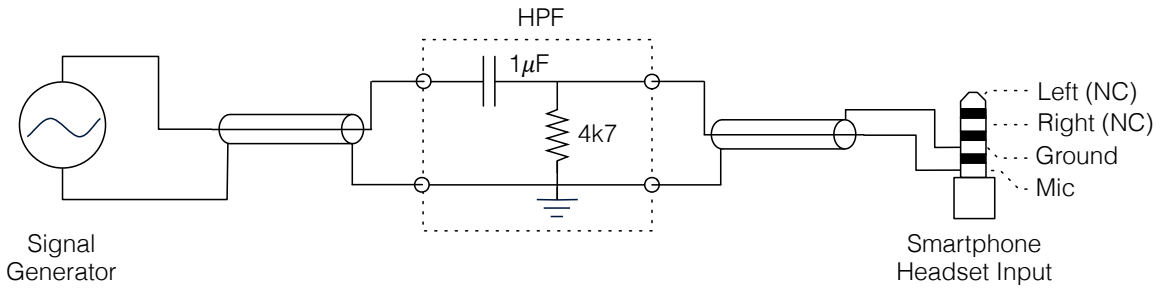


Figure 3.22: Cable used to measure the earplug audio input in several commercially available smartphones.

$$\mathrm{HPF}_{f_c} = \frac{1}{2\pi RC} = \frac{1}{2\pi \times 4k7 \times 1\mu} = 33.9 \text{ Hz} \tag{3.23}$$

The signal generator was implemented using a laptop computer equipped with an external soundcard (Edirol FA-66), with the sampling frequency set to 192 kHz to generate a ten second up-chirp with constant amplitude, starting at DC and ending at 96 kHz. This procedure was then repeated with the cable presented in Figure 3.22 interconnecting the output of the soundcard and the earplug input of the device under test. The resulting frequency response can be observed in Figure 3.23. Due to limitations in the hardware of the devices under test, the sampling frequency was set to the maximum value allowed, viz. 44.1 kHz.

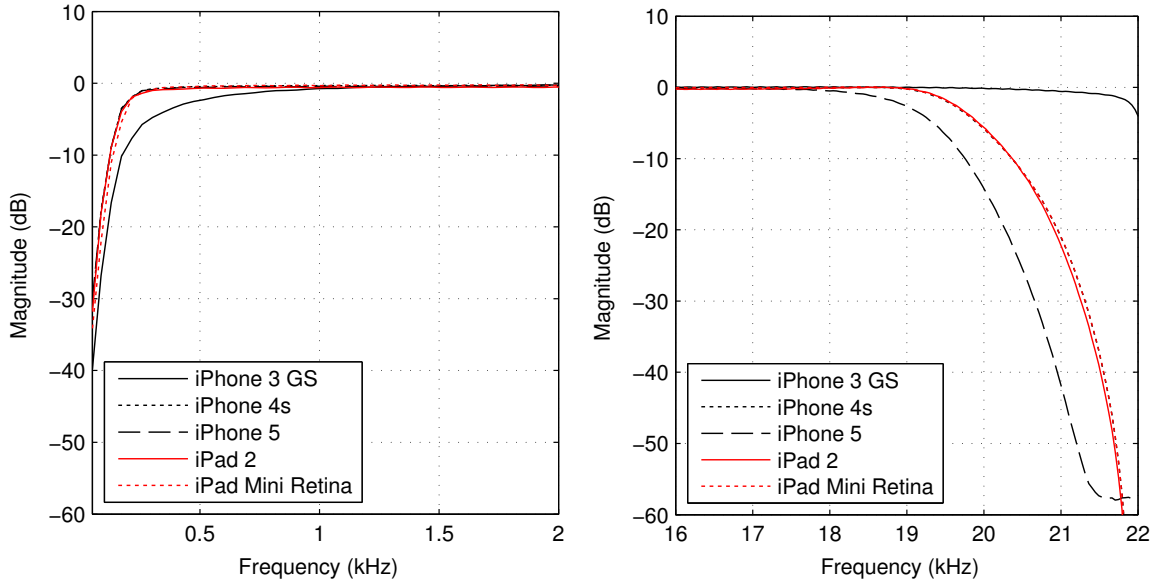


Figure 3.23: Audio input frequency response in several commercially available smartphones. **(a)** Low frequency response detail. **(b)** High frequency response detail.

Table 3.2 shows the results obtained for the following devices: iPhone 3 GS/4s/5 and iPad 2/Mini Retina.

Model	f_L @ -3 dB	f_H @ -3 dB
iPhone 3 GS	420 Hz	21.9 kHz
iPhone 4s	205 Hz	19.7 kHz
iPhone 5	187 Hz	19.1 kHz
iPad 2	195 Hz	19.7 kHz
iPad Mini Retina	205 Hz	19.7 kHz

Table 3.2: Smartphone headset audio input lower and upper cutoff frequencies.

3.4 Pulse Detection

In active ranging a deterministic signal is transmitted and should be detected to obtain its correspondent time of flight (ToF) measurement. To formulate the detection process some assumptions must be made, with the received signal $y[n]$ defined as:

$$y[n] = As[n - \tau] + w[n] \tag{3.24}$$

where A represents the signal amplification at the transmitter and amplitude attenuation due to propagation losses, $s[n - \tau]$ represents the line-of-sight (LoS) delayed version of a known deterministic signal $s[n]$, and $w[n]$ is AWGN with zero mean and σ^2 variance.

Additionally, it is important to detect the presence and position of the known signal $s[n]$ in the received signal $y[n]$ such as presented in Equation 3.24. In the presence of noise, it is necessary to decide if a signal is present or if the data contains only noise. This is a binary statistical problem that can be described by two hypotheses,

$$\mathrm{H}_0 \quad : \quad y[n] = w[n] \qquad n = 0, 1, 2, \ldots, N-1 \tag{3.25}$$

$$\mathrm{H}_1 \quad : \quad y[n] = w[n] + As[n - \tau] \qquad n = 0, 1, 2, \ldots, N-1 \tag{3.26}$$

where H_0 is known as the *null hypothesis*, H_1 is known as the *alternative hypothesis* and $w[n]$ is AWGN as defined before.

The *null hypothesis* H_0 states that the observed data consists of only noise. On the other hand, the *alternative hypothesis* H_1 states that the observed data consists of a deterministic signal plus noise. To make a decision it is necessary to formulate a rule based on a specific criterion to choose one between the two hypotheses. Figure 3.24 shows an example of this binary detection problem where it is possible to observe two probability density functions (PDF), one for noise and the other for signal plus noise. For a given criterion (in this case an energy threshold value) a decision can be taken, i.e. if energy is above the criterion value the hypotheses H_1 is selected and if the energy is below the criterion value the hypotheses H_0 is selected.

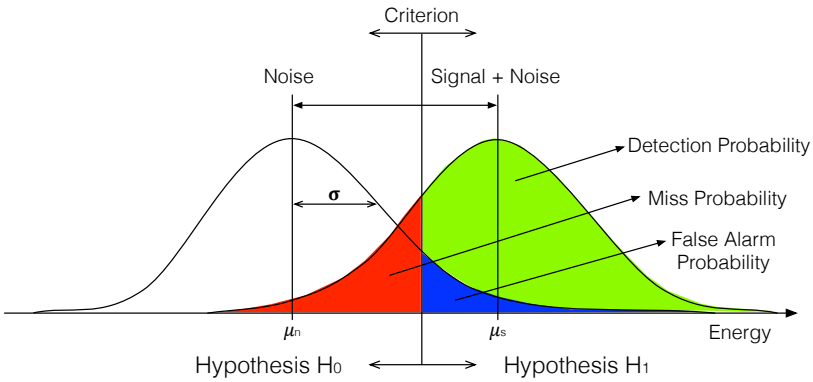


Figure 3.24: Example of signal detection in the presence of noise.

The area below the signal plus noise PDF is divided in three sub-areas that are related to the detection probability (H_1 is correctly selected, a signal is present), false alarm probability (H_1 is wrongly selected, only noise is present) and miss probability (H_0 is wrongly selected, a signal is present), for a given hypothesis test criterion.

3.4.1 Neyman-Pearson Detector

In applications such as radar and sonar the Neyman-Pearson (NP) optimality criterion is typically used to formulate the hypothesis test, in which the probability of detection is maximized such that the probability of false alarm remains below a specified level. When using the NP criterion the optimal decision rule derives from an optimal likelihood ratio test, to choose between H_0 and H_1 [86], cf. Equation 3.27.

$$L(\mathbf{y}) = \frac{p(\mathbf{y}; H_1)}{p(\mathbf{y}; H_0)} > \gamma, \tag{3.27}$$

where $\mathbf{y} = [y[0]\ y[1] \ldots y[N-1]]^T$. Using the PDF of the data under both hypotheses, we have

$$p(\mathbf{y}; H_1) = \frac{1}{(2\pi\sigma^2)^{\frac{N}{2}}} e^{-\frac{1}{2\sigma^2}\sum_{n=0}^{N-1}(y[n]-s[n])^2} \tag{3.28}$$

$$p(\mathbf{y}; H_0) = \frac{1}{(2\pi\sigma^2)^{\frac{N}{2}}} e^{-\frac{1}{2\sigma^2}\sum_{n=0}^{N-1} y^2[n]} \tag{3.29}$$

From the likelihood Equation 3.27 and using the PDF Equations 3.28 and 3.29 we can take the logarithm of both sides without changing the inequality,

$$\ln(L(\mathbf{y})) = \ln\left(\frac{p(\mathbf{y}; H_1)}{p(\mathbf{y}; H_0)} > \gamma\right) \tag{3.30}$$

$$= -\frac{1}{2\sigma^2}\left(\sum_{n=0}^{N-1}(y[n]-s[n])^2 - \sum_{n=0}^{N-1} y^2[n]\right) > \ln\gamma, \tag{3.31}$$

thus deciding for H_1 if

$$\frac{1}{\sigma^2}\sum_{n=0}^{N-1} y[n]s[n] - \frac{1}{2\sigma^2}\sum_{n=0}^{N-1} s^2[n] > \ln\gamma. \tag{3.32}$$

Incorporating the energy term into the threshold yields

$$T(\mathbf{y}) = \sum_{n=0}^{N-1} y[n]s[n] > \sigma^2 \ln \gamma + \frac{1}{2}\sum_{n=0}^{N-1} s^2[n], \tag{3.33}$$

and based on this new threshold value γ' we decide for H_1 if

$$T(\mathbf{y}) = \mathbf{y}^T\mathbf{s} = \sum_{n=0}^{N-1} y[n]s[n] > \gamma'. \tag{3.34}$$

The above statistical hypothesis test $T(\mathbf{y})$ is known as the NP-detector and is optimal for a deterministic signal in the presence of AWGN. It uses the upper bound of false-alarm probability as a constraint to obtain the γ' criterion. It is known in the literature as the replica-correlator and vulgarly implemented as a matched filter.

3.4.2 Matched Filter

A matched filter is a Linear Time Invariant (LTI) filter with a finite impulse response (FIR) that is perfectly matched to the signal that is being used ($s[n]$) in a flipped version of it. The decision is made by comparing the output of the matched filter with the threshold (γ') criterion [86].

The matched filter implementation of the Neyman-Pearson detector weights the samples with more energy more heavily than those of small energy and is well known by these two important properties:

1. Emphasizes frequency bands with higher signal power,

$$H(f) = S^*(f)e^{-j2\pi f(N-1)} \tag{3.35}$$

2. Maximizes the SNR at the FIR filter output,

$$\eta_{max} = \frac{\mathbf{s}^T\mathbf{s}}{\sigma^2} = \frac{\varepsilon}{\sigma^2} \tag{3.36}$$

where ε is the signal energy, $\varepsilon = \sum_{n=0}^{N-1} s^2[n]$.

Figure 3.25 illustrates the matched filter implementation for a deterministic signal detection in the presence of AWGN.

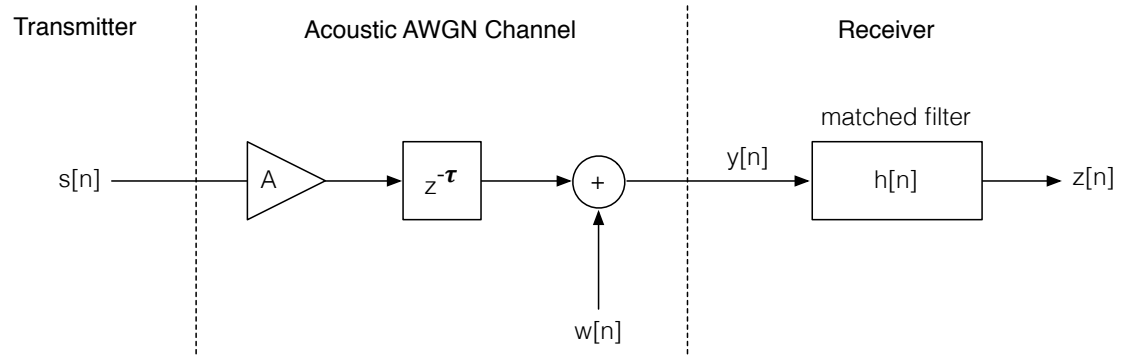


Figure 3.25: Detection with a matched filter.

The matched filter has a impulse response $h[n]$ given by

$$h[n] = s[N - n - 1], \quad n = 0, 1, \ldots, N - 1, \tag{3.37}$$

and its output $z[n]$ is given by

$$z[n] = \sum_{m=0}^{N-1} h[m]y[n - m]. \tag{3.38}$$

By combining Equations 3.37 and 3.38, the matched filter output can be computed by

$$z[n] = \sum_{k=0}^{N-1} s[N - k - 1]y[n - k]. \tag{3.39}$$

Its performance is given by its probability of detection (P_D), which can be obtained explicitly as a function of the probability of false alarm (P_{FA}) by

$$P_D = Pr\{T > \gamma'; H_1\} \tag{3.40}$$

$$= Q\left(Q^{-1}(P_{FA}) - \sqrt{\frac{\varepsilon}{\sigma^2}}\right). \tag{3.41}$$

The detection performance of a matched filter is better understood using the Receiver Operating

Characteristic (ROC) presented in Figure 3.26. From the figure it is possible to observe that an improvement in the detection performance (i.e. in the reduction of P_{FA}) can be obtained by an increase of the SNR at the output of the matched filter. In practice this can be achieved using two distinct ways: by increasing the signal amplitude or by increasing the signal duration in time. The shape of the signal itself does not affect the detection performance, being important only when other types of noise (coloured noise) are present [86].

By using a matched filter for signal detection it is possible to achieve a considerable processing gain due to its optimal hypotheses testing for a given probability of detection when AWGN is involved. For a constant amplitude pulse (i.e. rectangular window) the processing gain can be obtained by

$$PG = 10 \log_{10} N \tag{3.42}$$

where N is the pulse size in samples.

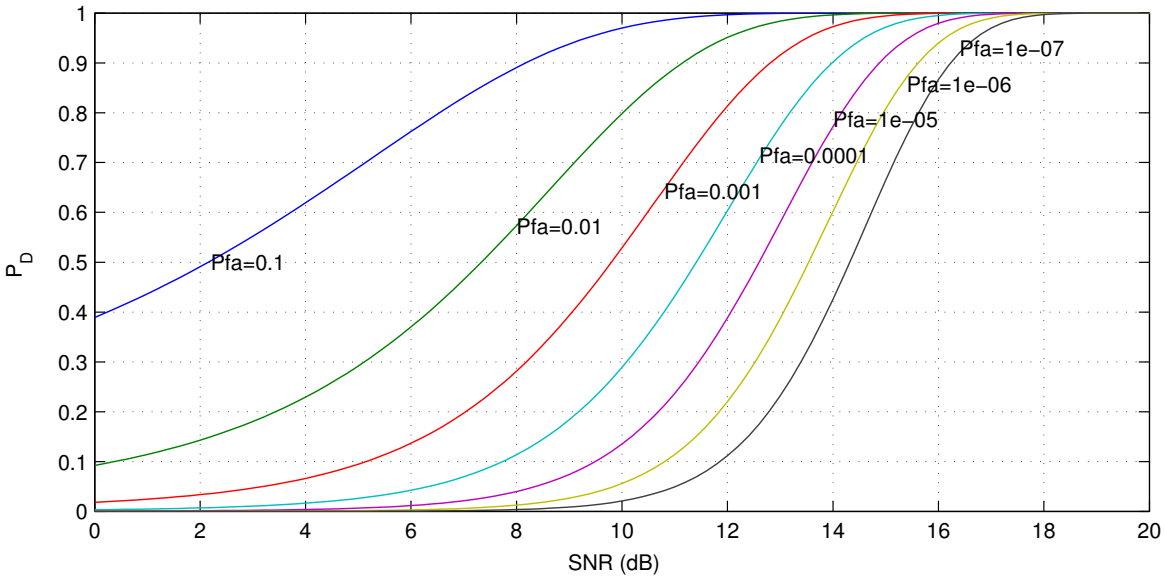


Figure 3.26: Detection performance of the matched filter.

3.5 Pulse Design

Signals with time and frequency diversity, e.g. chirps, are well known in radar and represent a case where both time and frequency are used to increase the probability of detection. Time bandwidth product (TBP) gives us the relation between the time duration of a signal and the range of frequencies (Bandwidth) necessary to its reconstruction. In radar, chirps with large TBP are used to obtain narrow compressed peaks and SNR maximization, resulting in signals with increased probability of detection, and also when Doppler tolerance is needed. By increasing TBP and using adequate weighting in the signal design it is possible to increase: the SNR, the

pulse compression (better time resolution) and the Doppler tolerance, which highly improves the probability of detection in static and dynamic positioning scenarios [87].

There are several constraints that we need to take into account in the pulse design process, of which we highlight:

1. Sound Propagation Constraints:
   - Use a link budget analysis to combine the transmitted power with the detection sensitivity to work in a specific range interval, cf. Appendix D.
   - Compensate the attenuation in the propagation of sound to bound the worst case scenario for the reference pulse.
2. Perceptual Impact:
   - Take advantage of the human hearing sensitivity by modulating in amplitude the pulses using the proposed perceptual filter;
   - Limit the pulse duration to 10 ms to take advantage of the temporal integration of the hearing system thus reducing the loudness sensation by more than 10 dB;
   - Limit the pulse bandwidth to a small number of critical bands to reduce the loudness sensation.
3. Hardware Limitations:
   - Use the available bandwidth of the audio input of a conventional smartphone.
   - Consider the processing capabilities of the smartphone, e.g. availability of MAC operations and optimized DSP frameworks.
4. Matched Filter Performance:
   - Limited probability of detection for a given probability of false alarm (ROC curve).

3.5.1 Approach 1: Time Domain Weighting

In this approach we will focus on time domain weighting to modulate the transmitted pulse in amplitude in order to design a non-invasive audio chirp pulse. This pulse must be carefully designed to avoid spectral spreading that normally appears when fast transients occur.

3.5.1.1 Design Procedure

This approach leads to a significant reduction in the transmitted power and therefore a considerable SNR reduction at the output of the matched filter. To maximize the transmitted power, a combined weighting window is used to modulate the transmitted chirp pulse in amplitude using two distinct parts. Both parts can be generated using a tapered cosine window (also known as Tukey window). A Tukey window is a rectangular window with the first and last $r/2$ fractions of the samples equal

to parts of a cosine, were r is a real number between 0 and 1. Note that a Hanning window can be seen as a particular case of the Tukey window, when $r = 1$. In this case, the proposed window uses the left half of a Hanning window, concatenated with the second half part of a Tukey window with $r = 0.05$, for a chirp pulse duration of 10 ms, cf. Figure 3.27.

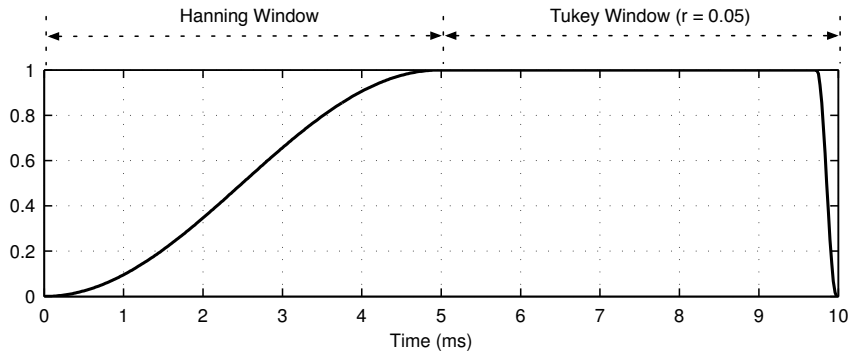


Figure 3.27: Proposed combined window.

Figure 3.28 compares other weighting windows with the proposed using different representations, from which are extracted several figures of merit for better evaluation. Two of them, namely the Compression Ratio[2] (CR) and the Peak-Sidelobe-Level[3] (PSL), evaluate the pulse compression gain. For the proposed weighting pulse (e), CR presents a value of 0.40 ms. This value was measured looking for the time interval where the autocorrelation peak drops 6 dB. The PSL value represents the difference between the autocorrelation peak and the nearest side-lobe peak and was measured with 16.6 dB.

The proposed window enables the increase of the Peak Level (PL) energy at the output of the matched filter by more than 5 dB at the cost of a significative reduction of the PSL (Figure 3.28), but with the advantage of an increased pulse compression ratio, cf. Table 3.3.

Chirp Pulse	Weighting Window	B (kHz)	TBP	CR (ms)	PSL (dB)	PL (dB)
(a)	Rectangular	2	60	0.60	14.1	0.0
(b)	Hanning	2	60	1.34	46.0	-8.5
(c)	Rectangular	4	120	0.30	13.6	0.0
(d)	Hanning	4	120	0.62	46.4	-8.5
(e)	Combined	4	120	0.40	16.6	-3.5

Table 3.3: Figures of merit of the chirp pulses presented in Figure 3.28.

[2]CR is measured by looking for the time in which the main autocorrelation peak is above a -6 dB threshold line.
[3]PSL was measured as the interval, in dB, between the main autocorrelation peak and the nearest sidelobe.

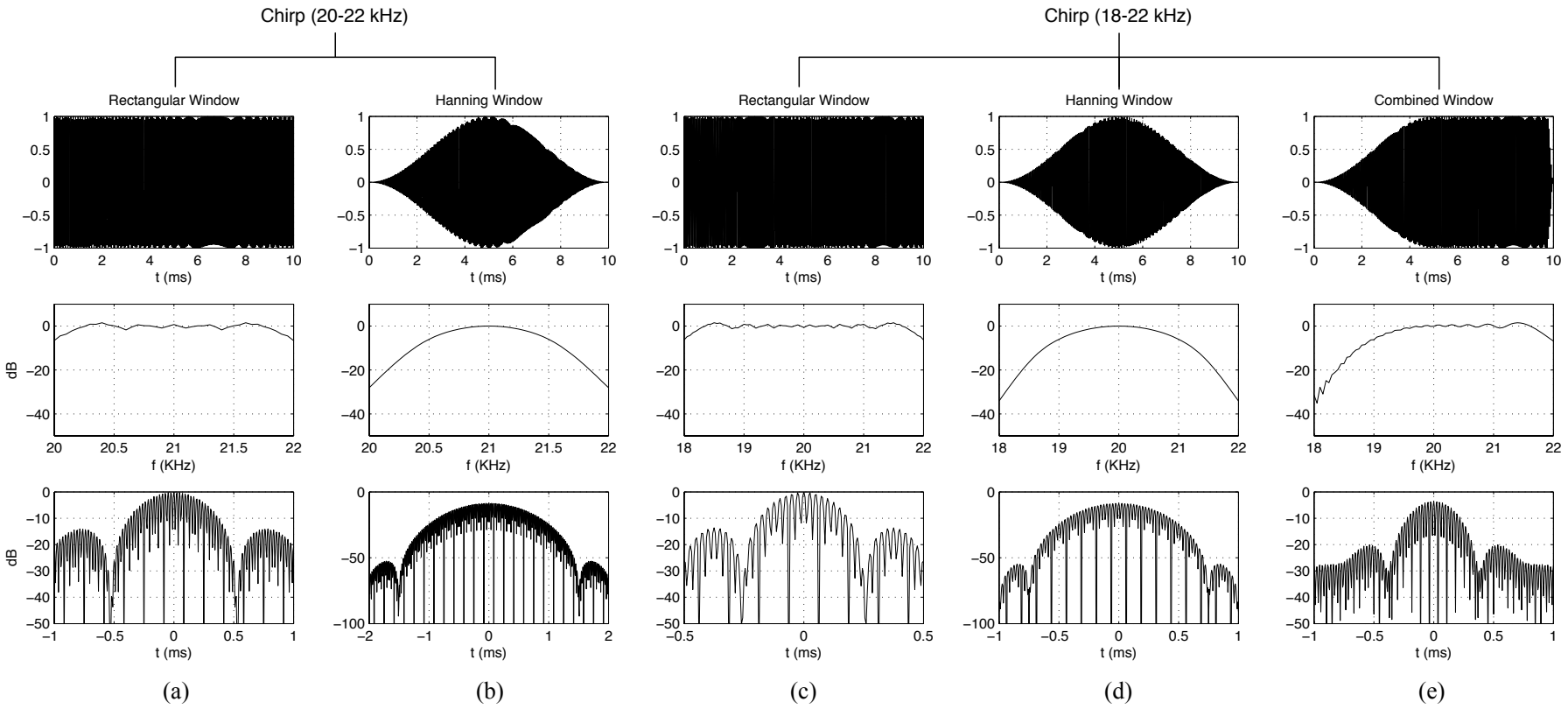


Figure 3.28: Chirp pulse design using time domain weighting. (a) and (b) columns correspond to a chirp pulses with frequencies raising from 20 to 22 kHz; (c), (d) and (e) columns are related to chirp pulses with frequencies raising from 18 kHz to 22 kHz. First line of plots represents the weighted pulses in the time domain, second line represents its frequency content and the third line represents its autocorrelation function in time around the central peak.

3.5.2 Approach 2: Perceptual Filtering

The perceptual filtering approach uses the filter proposed in section 3.2.3.1 to reduce the perceptual impact of the pulses. This filter is based on the results by Suzuki and Takeshima in 2004 [10] where they estimate the threshold of hearing contour based on twelve contemporary loudness studies. Our design approach starts with the design of a reference short time pulse of 10 ms that will be used as a yardstick in the design of other pulses for evaluation.

3.5.2.1 Design Procedure

As several pulses will be proposed at this stage we have opted for designing a reference pulse. The main goal is to design distinct pulses with two hard constraints: same energy (E) and same bandwidth (B). The energy value of the reference pulse will be used to generate pulses with the same energy but with different durations (T_i).

To obtain the reference pulse we have used the following procedure:

(1) Set a fixed duration for the reference pulse, $T_0 = 10$ ms

(2) Set a fixed bandwidth for the reference pulse, B =8 kHz with frequencies 16–24 kHz

(3) Generate an up-chirp pulse using a rectangular time domain window based on the constraints defined in (1) and (2)

(4) Filter the pulse generated in (3) with the perceptual FIR filter obtained in section 3.2.3.1.

(5) Gate the pulse on and off with two raised cosine ramps of 1 ms duration to guarantee smooth transients and thus turn the onsets and offsets inaudible.

(6) Compute the energy (E) and the power (P) of the generated pulse.

$$E = \sum_{n=0}^{N-1} x^2[n] \tag{3.43}$$

$$P = \frac{1}{N} \sum_{n=0}^{N-1} x^2[n] \tag{3.44}$$

Figure 3.29 presents the autocorrelation function of the reference pulse generated using the procedure introduced before for a given energy value (in this case $E = 80$) which resulted in a central peak magnitude of approximately 38 dB. In the pulse design process several figures of merit will be used to evaluate the pulse detection performance when using a matched filter. All considered pulses will be generated with the same energy for a better and fair comparison of their

figures of merit. Moreover the pulse bandwidths will be forced to be the same before filtering with the perceptual filter obtained using the data estimated by Suzuki and Takeshima [10].

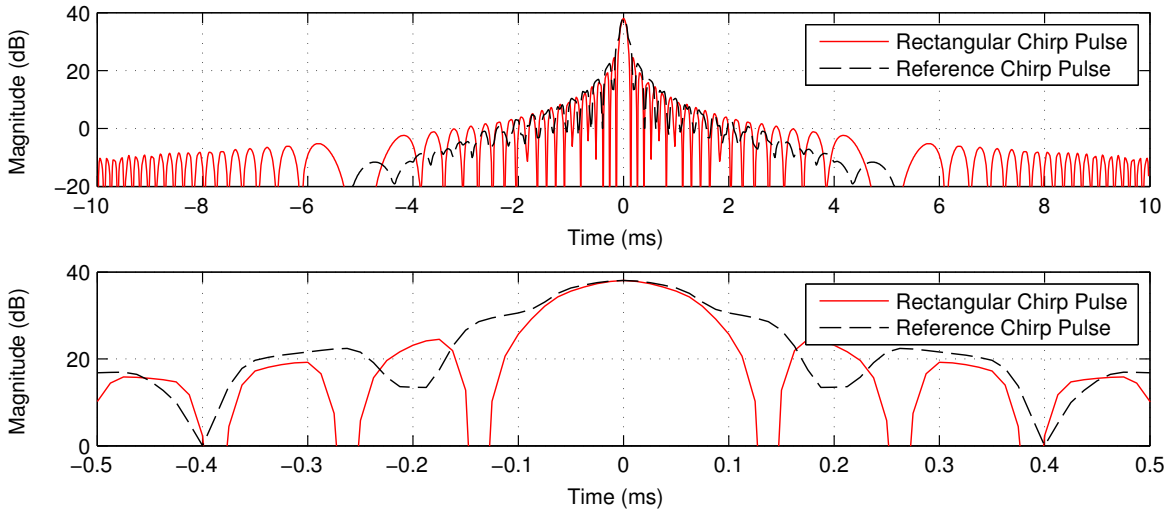


Figure 3.29: Envelope of the autocorrelation function of a rectangular window up-chirp pulse and the reference perceptual filtered pulse.

Based on this reference pulse, several other pulses will be computed to match the same energy E but with different pulse durations. The selected durations were: 10 ms (reference duration), 20 ms, 40 ms, 80 ms, 160 ms, 320 ms, 640 ms and 1280 ms. Moreover, three types of up-chirp pulses will be considered (linear, logarithmic and quadratic) and different duration BPSK pulses with the same bandwidth which results in Kasami sequences with different lengths. These four types of pulses combined with the eight different durations resulted in 32 distinct pulses for comparison (reference pulse included). The following procedure was used to generate these pulses:

(1) Select a pulse type among the following: up-chirp (linear, logarithmic or quadratic) or a BPSK modulated Kasami sequence.

(2) Select a specific duration (T_i) for the pulse from these durations: 10 ms(reference duration), 20 ms, 40 ms, 80 ms, 160 ms, 320 ms, 640 ms and 1280 ms.

(3) Set the same fixed bandwidth (B) used in the reference pulse. If a BPSK pulse was selected in (1), a Kasami sequence with length N should be computed based on the available Bandwidth, i.e. B=8 kHz, such as

$$N = \frac{B}{2} \times T_i \tag{3.45}$$

which will result in one of the following Kasami sequence lengths: 20 bit, 40 bit, 80 bit, 160 bit, 320 bit, 640 bit, 1280 bit and 2560 bit, for the pulse durations previously introduced in (2).

(4) Generate the pulse using a rectangular time domain window based on the constraints defined in (1), (2) and (3).

(5) Remove all frequency components below 16 kHz and above 24 kHz in the BPSK signals using a bandpass FIR filter with a 60 dB rejection in the stopband and bandpass cutoff frequencies at 16 kHz and 24 kHz.

(6) Filter the pulse generated in (5) with the perceptual FIR filter obtained in section 3.2.3.1.

(7) Gate the pulse on and off with two raised cosine ramps of 1 ms duration to guarantee smooth transients and thus turn the onsets and offsets inaudible.

(8) Adjust the gain of the pulse obtained in (7) to match the same energy E of the reference pulse.

Figure 3.30 presents a block diagram of the pulse design procedure previously introduced.

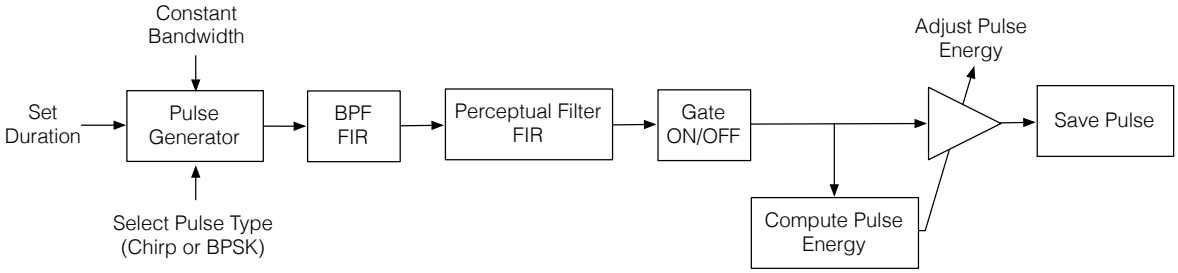


Figure 3.30: Block diagram of the pulse design procedure using the perceptual filter.

As the pulse energy and bandwidth are constant, to evaluate the pulses quantitatively we will consider three important figures of merit of the autocorrelation function, namely: Peak-Sidelobe Level (PSL), pulse Compression Ratio (CR) and Peak-to-Average Ratio (PAR). These figures of merit will be used to evaluate the pulse detection performance after the matched filter.

Figure 3.31 shows the autocorrelation functions around the central peak for all the proposed up-chirp and BPSK-based pulses. From the figures it is possible to observe that the main peak of the autocorrelation function has the same value for all the proposed pulses. This indicates that all the pulses have the same energy, although they have different durations.

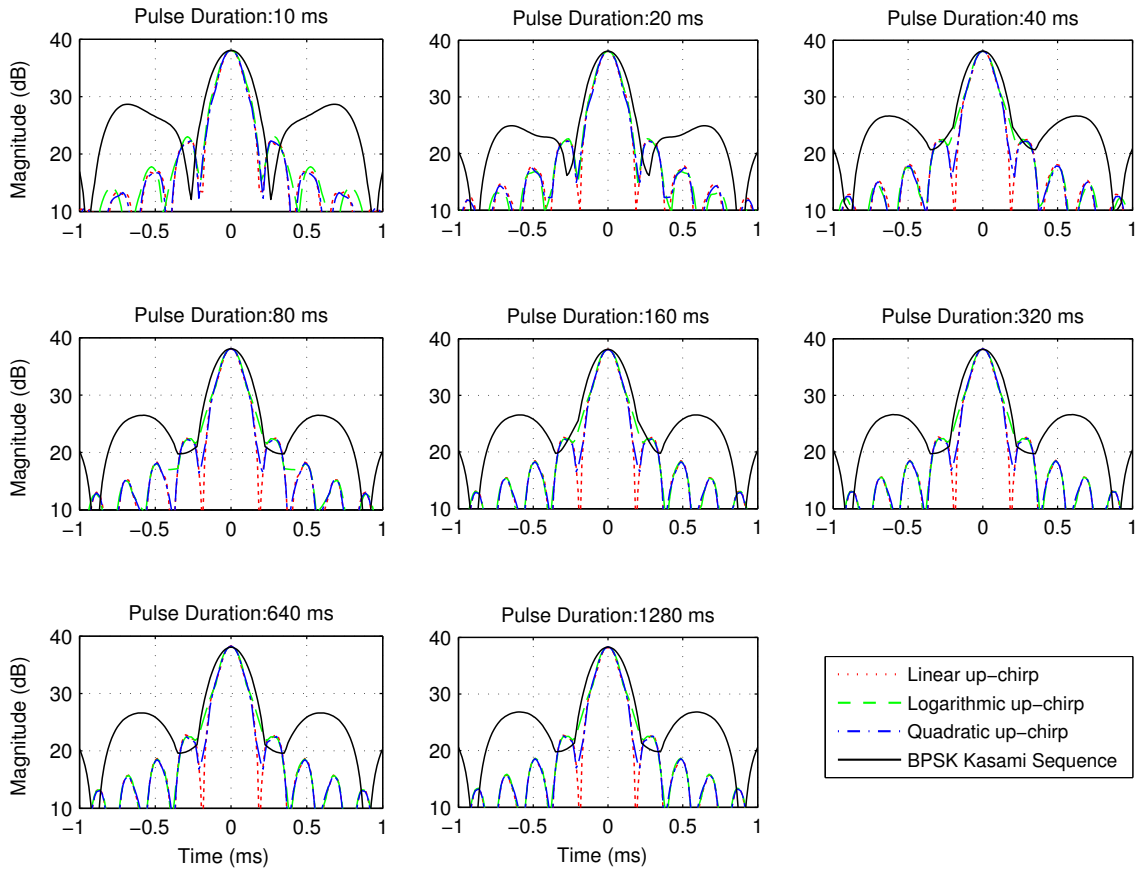
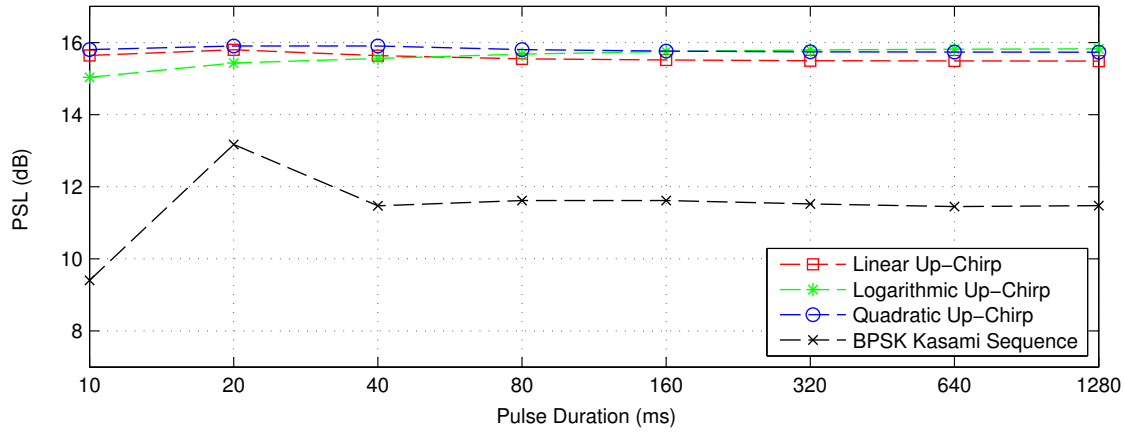


Figure 3.31: Envelope of the autocorrelation functions around central peak for different type pulses and durations.

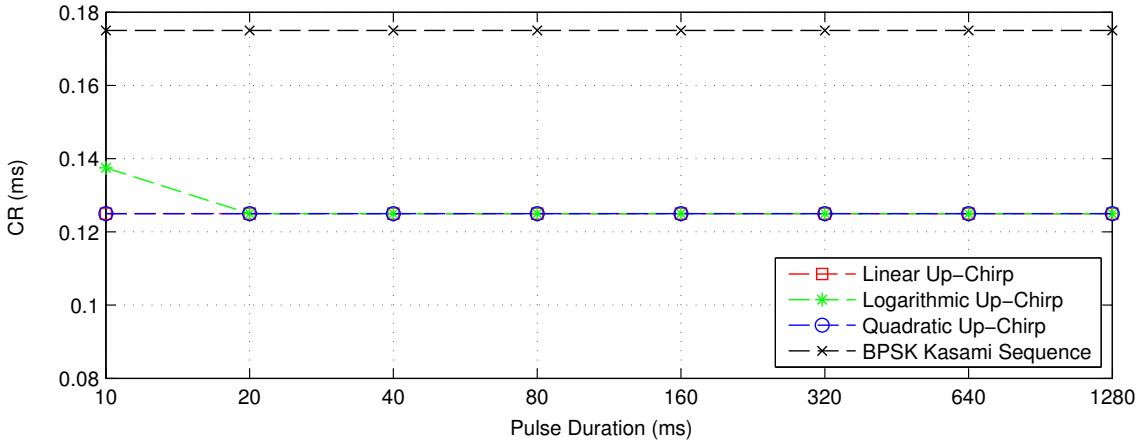
All the chirp-based pulses present approximately the same PSL, which is influenced by the time domain envelope of the pulse after perceptual filtering. In the case of high accuracy measurements it is important to have a higher PSL to reduce the jitter in peak detection. Moreover, if AWGN is considered, it is important to detect the main peak at the output of the matched filter rather than the side lobe position in time. In this case a higher PSL leads to a better main peak discrimination thus improving the peak detector performance. The use of chirp-based pulses takes advantage of this figure of merit, cf. Figure 3.32a.

Figure 3.32b depicts the pulse CR after the matched filter. The CR figure of merit of the autocorrelation function remains small (below 0.14 ms) and constant for the chirp-based pulses. For the BPSK-based pulses the CR degrades, settling to a constant (approximately 0.18 ms) value for all distinct pulse durations. This can be justified by the well known pulse compression property that chirps present when used with matched filters, i.e. the compression ratio is only dependent of the available bandwidth B and not in the waveform shape or duration [87].

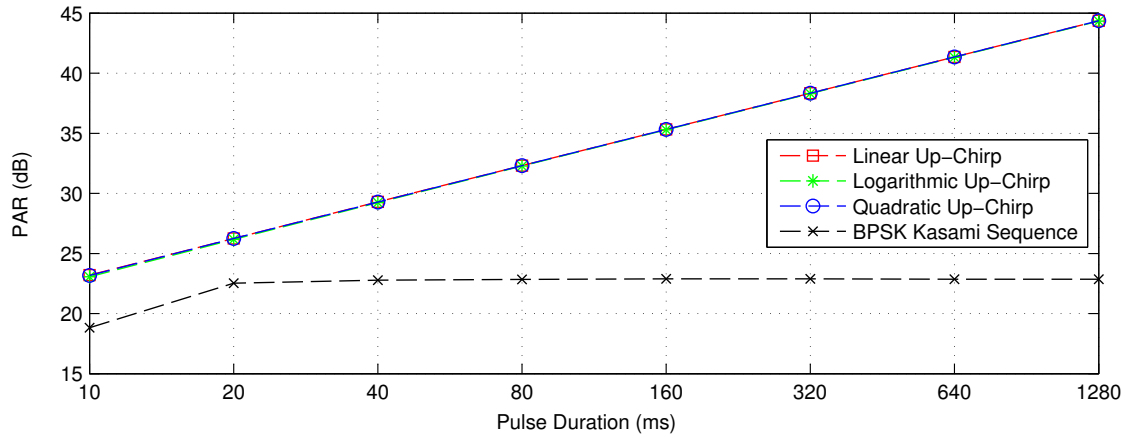
Figure 3.32c presents the PAR of the autocorrelation functions. The BPSK-based pulses tend to a constant PAR when the pulse duration is greater that 20 ms. On the other hand all the



(a) Peak to Sidelobe Level (PSL).



(b) Compression Ratio (CR).



(c) Peak to Average Ratio (PAR).

Figure 3.32: Figures of merit for the autocorrelation functions presented in Figure 3.31.

chirp-based pulses present a similar and approximately linear increase of PAR with the pulse duration increase. This figure of merit shows us that the chirp based signals always present better PAR values. As we are considering signals with the same energy, the peak discrimination (also when AWGN is included) is always more pronounced with chirp-based pulses, which enhances the performance of the peak detector.

The design procedure exchanged time for power maintaining the pulse energy at the output of the matched filter. By using a perceptual filter, if a user is exposed to the reference pulse (10 ms duration pulses) and its average power is calibrated to have reduced impact, all the pulses proposed here should present an equal or smaller perceptual impact. Figure 3.33 shows the average power (AP) for each designed pulse and it is possible to observe the reduction of the AP with the increase of the pulse duration.

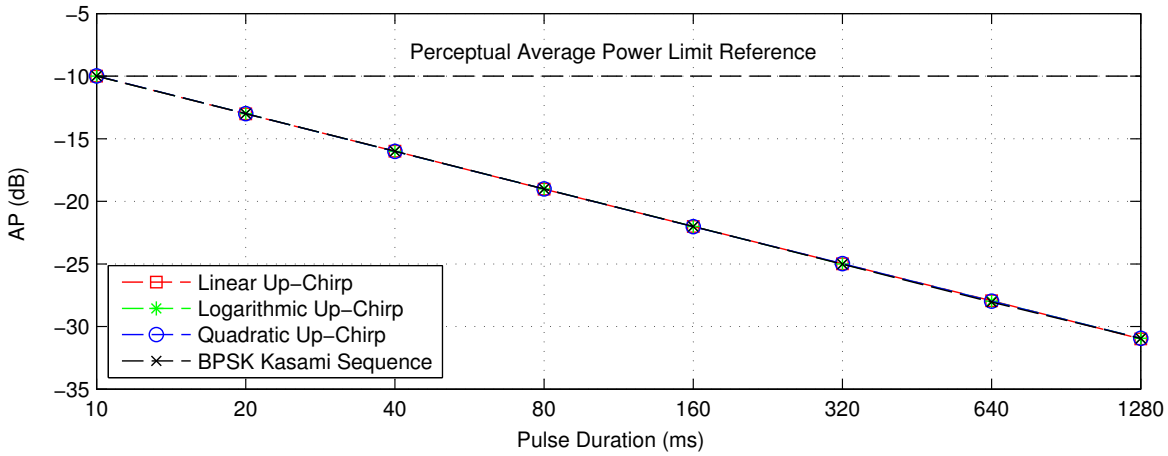


Figure 3.33: Average Power (AP) of the proposed pulses.

This gives us margin to increase the power at the transmitter, for the pulses with duration greater than 10 ms, to approximate the perceptual average power limit line. This increase in average power maintains the perceptual impact of the pulse reduced, but increases the pulse energy, which increases the probability of detection, cf. Figure 3.26. At the end of this chapter we evaluate the perceptual impact of some of the proposed pulses with subjects in a soundproof room.

3.5.3 Doppler Tolerance

To evaluate the Doppler resilience of the proposed signals their ambiguity function will be used. The ambiguity function is a well known radar tool that represents the output of the matched filter in the presence of a time delay variation τ versus a Doppler shift (f) caused by a moving receiver with velocity v. In our case the ambiguity function will be used to evaluate the Doppler shift in range measurements between a static acoustic transmitter and a moving receiver. It is important to

observe that the ambiguity function at $(\tau, f) = (0, 0)$ is equal to the matched filter output when it is perfectly matched to the transmitted signal, i.e. when the receiver is also static. The ambiguity function at nonzero τ and f give us information about the change of the matched filter output when a receiver is moving. Equation 3.46 represents the continuous time version of the narrowband ambiguity function for a given complex baseband pulse $x(t)$.

$$\chi(\tau, f) = \int_{-\infty}^{\infty} x(t)x^*(t-\tau)e^{i2\pi ft}\, dt \tag{3.46}$$

where * represents the complex conjugate. Note that for zero Doppler shift ($f = 0$) this simplifies to the autocorrelation function of $x(t)$. In our application case the Doppler shift (f) is related with the speed of sound (c) and the velocity v of the receiver, resulting in a shifted received frequency (f_r) as a function of the original transmitted frequency (f_t) :

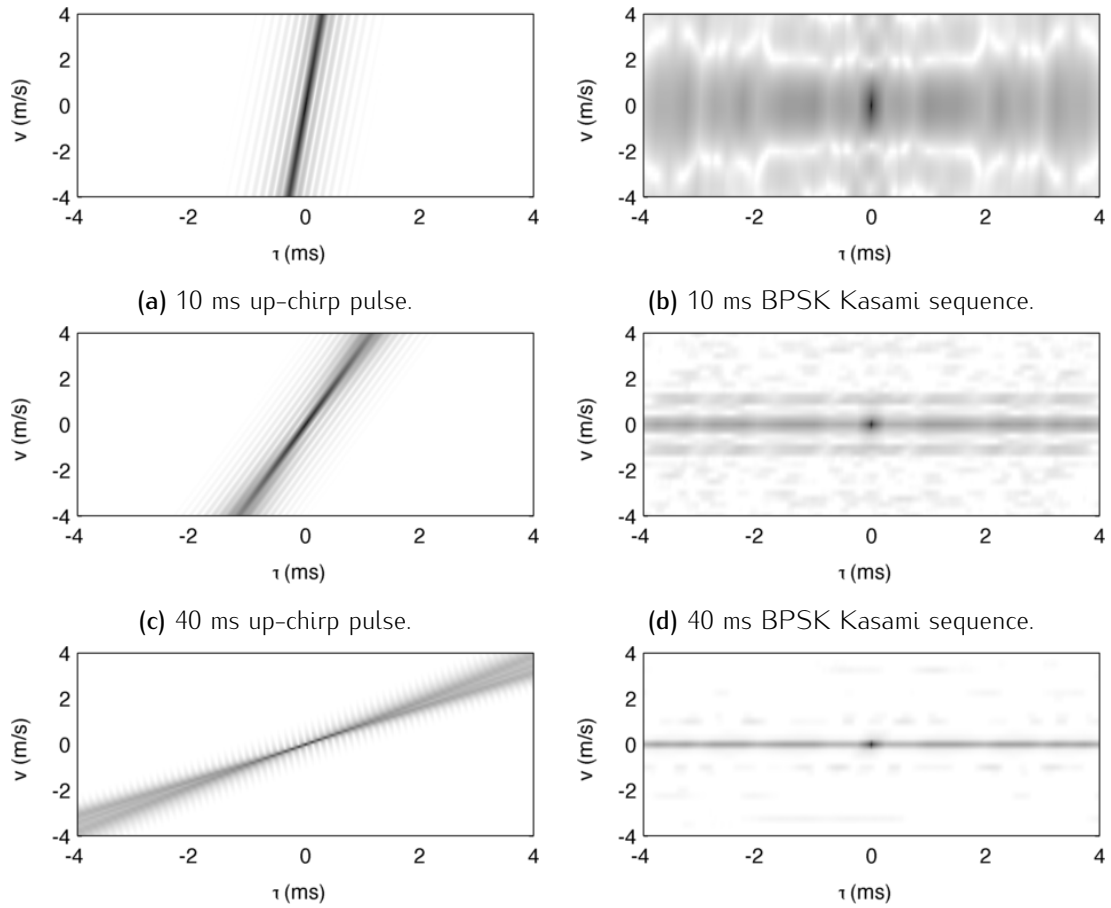
$$f_r = f_t \left(\frac{1 + v/c}{1 - v/c} \right). \tag{3.47}$$

The Doppler shift (f) is then obtained by

$$f = f_r - f_t = 2v \frac{f_t}{(c - v)}. \tag{3.48}$$

In our case the ambiguity function will be generated for Doppler shifts that appear when a receiver is moving with velocities up to 4 m/s. Note that we are interested in pedestrian tracking applications, so this velocity range is enough. We consider a time delay (τ) interval of 4 ms, which results in a distance estimation error of approximately 1.32 m (considering a sound speed of 343 m/s).

Figures 3.34a, 3.34c and 3.34e present the ambiguity functions for the linear up-chirp pulse with different durations as a function of time delay τ and velocity change v. From the figures, it is possible to observe that a velocity change in the receiver can still lead to a peak at the output of the matched filter although at the cost of a time-delay τ estimation error. Another observation is that when the pulse duration increases for a fixed bandwidth the ambiguity function suffers from a rotation that increases the τ error thus resulting in an increase of the ranging error. For example, for an up-chirp pulse of 160 ms if the receiver is moving at approximately 2 m/s the resulting ranging error is approximately 1 m.



(a) 10 ms up-chirp pulse.

(b) 10 ms BPSK Kasami sequence.

(c) 40 ms up-chirp pulse.

(d) 40 ms BPSK Kasami sequence.

(e) 160 ms up-chirp pulse.

(f) 160 ms BPSK Kasami sequence.

Figure 3.34: Ambiguity functions for the distinct pulses proposed. (a), (c) and (e) Linear up-chirps with different durations and same bandwidth. (b), (d) and (f) BPSK-based pulses with different Kasami sequences, different durations and same bandwidth.

Figures 3.34b, 3.34d and 3.34f present the ambiguity functions for the BPSK-based pulses with the same duration but different Kasami sequence lengths, as a function of time delay τ and velocity change v. As expected, when the pulse duration increases the number of bits of the Kasami sequence must increase to maintain the pulse bandwidth. This leads to a more compressed spike in the centre of the ambiguity function. Considering the 10 ms BPSK pulse (Figure 3.34b), for speeds greater than 0.5 m/s the peak at the output of the matched filter completely disappears, turning this pulse useless when the receiver velocity is greater that 0.5 m/s. Moreover, the BPSK pulses present poor figures of merit when compared to the chirp-based signals, viz. the lowest PSL and PAR values and the biggest CR values lead to worst performance in peak detection when using a matched filter.

3.6 Perceptual Evaluation

In this section we describe the experimental procedure used to perform the perceptual evaluation. The goal is to evaluate, in a controlled acoustic environment, the perceptual impact of the admittedly non-invasive pulses designed in the previous section.

The room used for the perceptual evaluation is the soundproof room available at the Speech, Language and Hearing Laboratory (SLHlab) in the School of Health Sciences, University of Aveiro (ESSUA). The use of this room guaranteed a background noise level below 40 dB (SPL) and avoided the presence of other noise sources (e.g. impulsive and continuous noise sources) that normally appear in typical rooms.

The perceptual evaluation was based on the ITU-R 5-point continuous rating scale [88] and was performed based on feedback obtained from the subjects for each tested pulse, cf. Table 3.4.

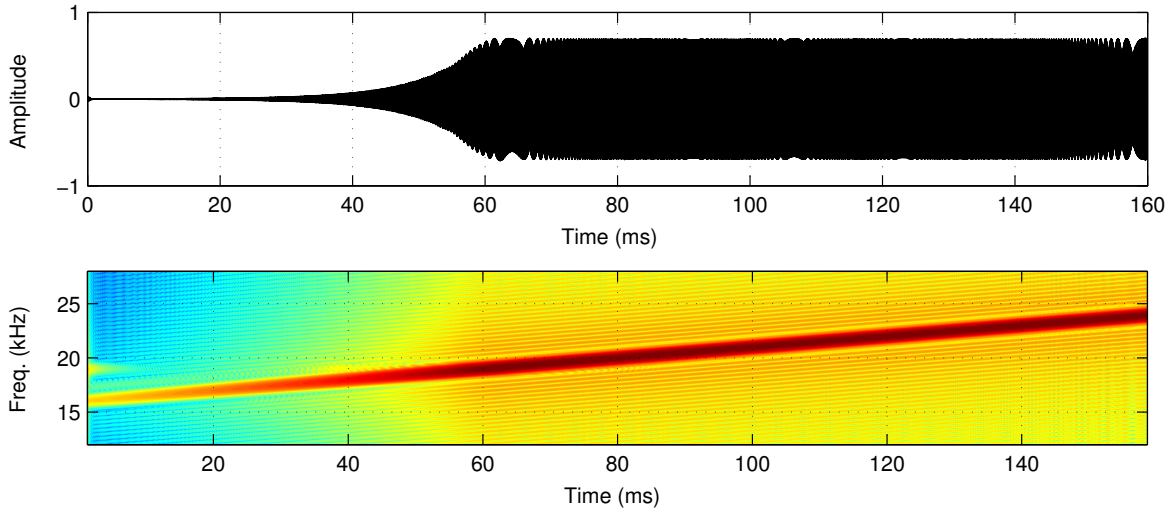
Classification	Qualitative Evaluation	Grade
Invasive	Very annoying	5.0
	Annoying	4.0
	Slightly annoying	3.0
Non-Invasive	Perceptible, but not annoying	2.0
	Imperceptible	1.0

Table 3.4: ITU-R 5-point continuous rating scale used in the perceptual evaluation form [88].

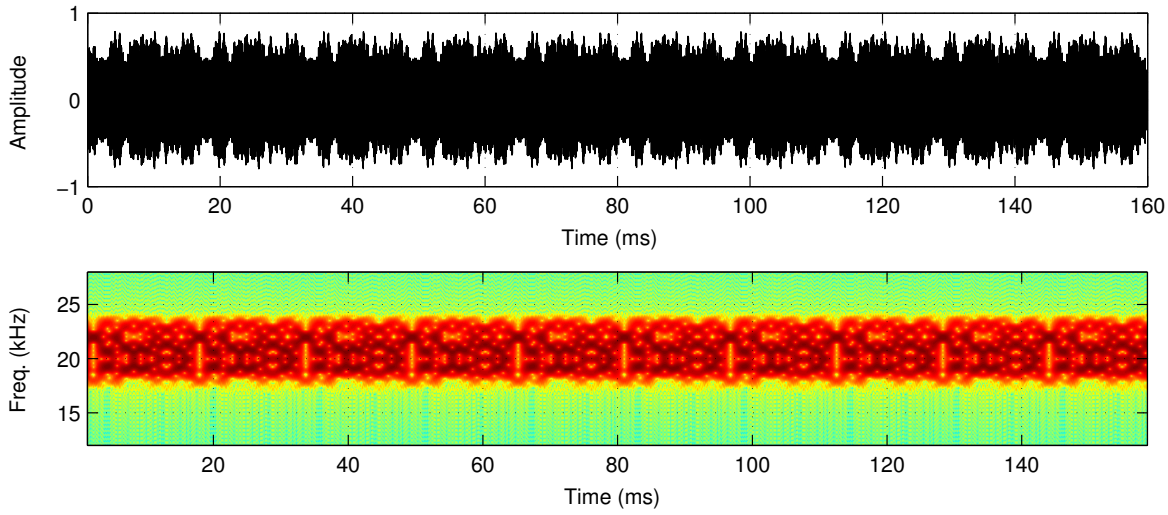
3.6.1 Experimental Procedure

We have recruited 10 male and 5 female naïve subjects with ages in the interval 16–51 years old. The mean age was 28.3 years (standard deviation 9.8 years). None of the listeners has reported hearing problems and they have not received financial compensation for their participation.

The pulses used as stimulus are the ones proposed in Section 3.5 for the following durations: 10 ms, 40 ms and 160 ms. Figure 3.35 shows the time domain and spectrogram representations of two example pulses used in the perceptual evaluation experiments.



(a) Linear up-chirp pulse with 160 ms of duration.



(b) BPSK pulse with 160 ms of duration.

Figure 3.35: Time domain and spectrogram representations of two of the pulses used in the perceptual evaluation. Both pulses were designed to have the same energy and bandwidth.

3.6.1.1 Experimental Design

Each subject has been submitted to a set of seven high frequency (with frequencies in the range from 16 kHz to 24 kHz) pulses for qualitative evaluation. The test method used for evaluation is the method of limits [88]. This method looks for the threshold estimation on the basis of the reversal, which is when the subjects change their response. In our case we will consider a change (based on the ITU-R 5-point continuous rating scale) from "Perceptible, but not annoying" to "Slightly annoying", i.e. when the subject evaluates quantitatively with a grade greater than 2.5.

The experiment starts by listening to each pulse from a loudness reference value of 55 dB (SPL), with a repetition frequency of 1 Hz. After the transmission of 15 equal pulses, i.e. 15 s of

pulse exposure, the user is asked to fill an evaluation form (Appendix B), using the scale previously introduced in Table 3.4. After 15 s a new set of the same pulse with a 5 dB (SPL) increase is evaluated. This procedure is repeated until a maximum of 75 dB (SPL) is achieved. This 75 dB (SPL) value was selected based on the recommended exposure limits for sound and ultrasound introduced in Table 3.1.

Figure 3.36 presents a flowchart of the experimental procedure. This procedure takes 30 s per iteration. Considering that we begin with 55 dB (SPL) the procedure ends in 5 rounds thus resulting in a total duration of 2.5 minutes per pulse evaluation. As we want to evaluate seven pulses, the overall duration of the experiment is approximately 20 min per subject group.

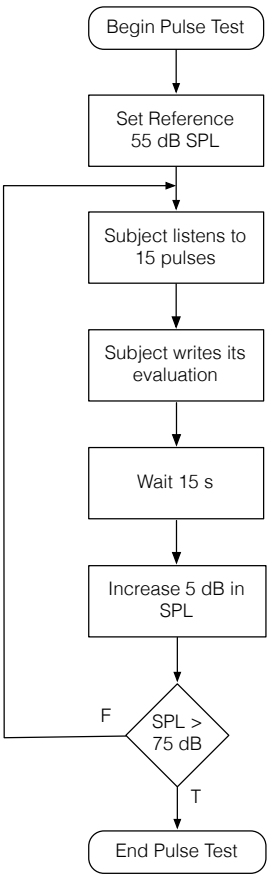
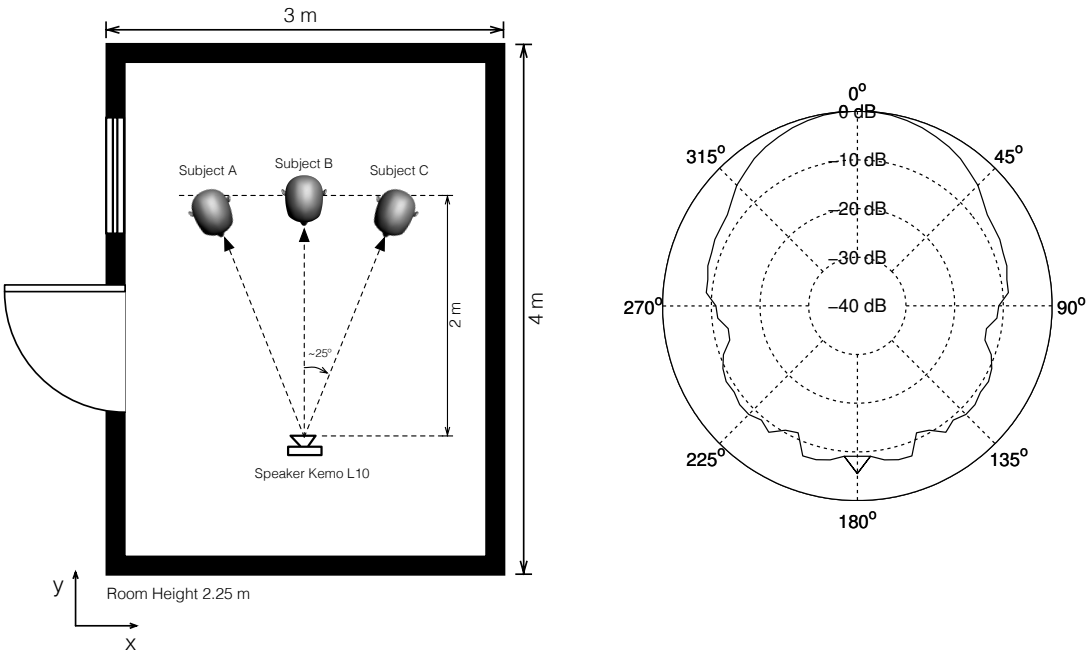


Figure 3.36: Flowchart of the perceptual evaluation experimental procedure.

3.6.1.2 Experimental Setup

The experiments were performed in the soundproof room of the SLHlab/ESSUA. The loudspeaker used in the experiment was a Kemo L10 piezo-tweeter, that is described in more detail in Section 4.4.1. Figure 3.37a shows the floor plan of the soundproof room with the experimental setup described. The Kemo L10 loudspeaker was placed at an approximate distance of 2 m from the subjects' heads in LoS. For angles below 30 degrees the loudspeaker attenuates less than 3 dB,

cf. Figure 3.37b. This allowed us to perform each experiment with three subjects at a time (Figure 3.37a), using the evaluation form presented in Appendix B. Both the loudspeaker and the subjects' head were placed at the same height of 1.5 m.



(a) Floor plan of the SLHlab/UA soundproof room. **(b)** Kemo L10 radiation pattern at 20 kHz.

Figure 3.37: Experimental Setup.

The setup presented in Figure 3.38 was used for calibration of the experiment. A Cirrus Research sound level meter, model CR:171B, was used to measure the SPL at a distance from the loudspeaker of 2 m and height of 1.5 m. The level meter was set to measure SPL using a flat frequency response, i.e. without weighting.

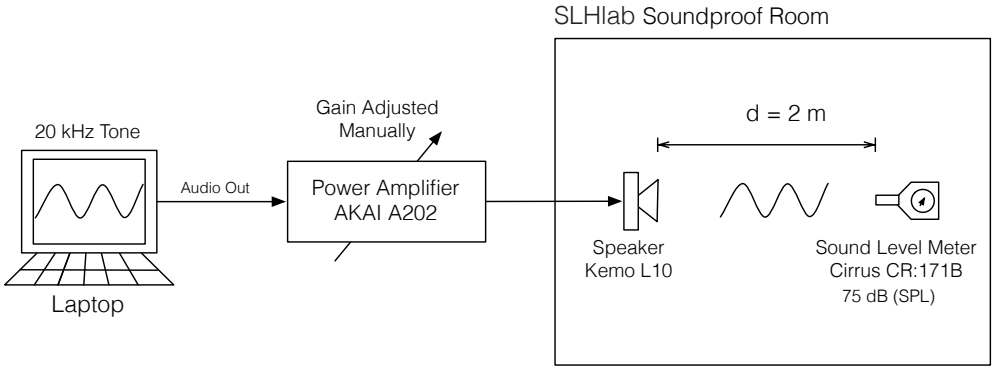


Figure 3.38: Calibration procedure.

In the calibration process, a continuous sine wave with a frequency of 20 kHz with normalized amplitude ±1.0 was generated with the Audacity application in the laptop using a sampling frequency of 96 kHz. This is the maximum amplitude in the digital path, just before clipping in the D/A conversion. The gain of the power amplifier AKAI A202 was then adjusted to obtain 75 dB (SPL) measured in the sound level meter.

3.6.2 Evaluation with Subjects

In the perceptual evaluation experiment (Figure 3.39) the gain of the power amplifier was constant. The 5 dB steps were updated in the analog path of the signal by acting in the output gain of the laptop soundcard using steps of 5 dB. The soundcard used was the built-in soundcard of a Macbook Pro (late 2007) running with a sampling frequency of 96 kHz. The background noise inside the soundproof room with the door closed was measured with the sound level meter and showed values below 40 dB (SPL).

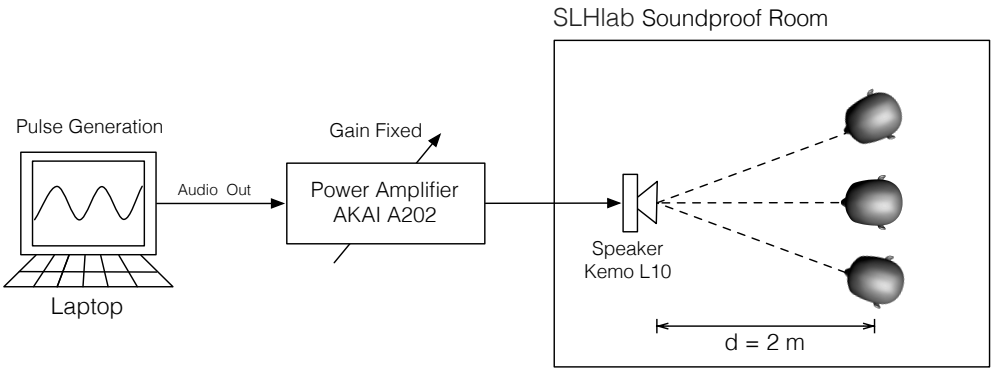


Figure 3.39: Experimental setup used in the perceptual evaluation.

The pulses were presented randomly to the subjects in seven distinct auditions, one for each pulse, cf. Table 3.5. For each audition, five rounds were calibrated to obtain the following SPL at a distance of 2 m from the loudspeaker: 55 dB, 60 dB, 65 dB, 70 dB and 75 dB.

Appendix B presents the evaluation form used by the users to perform the perceptual evaluation. The pulses were designed to have the same bandwidth and energy. This leads to different average power for each pulse, cf. Figure 3.33. To obtain the same SPL for all tested pulses, the amplitude of each pulse was previously adjusted with the gains presented in Table 3.5. These gains can be seen as an improvement of the detection probability at the output of the matched filter as discussed in Section 3.4.2. The subject answers are reported in Appendix C and will be analysed and discussed in detail in Section 6.1.

Audition Order	Pulse Design Approach	Pulse Type	Pulse Duration	Gain
1	Perceptual Filtering	Up-chirp	10 ms	1 dB
2		Up-chirp	160 ms	14 dB
3		BPSK	40 ms	7 dB
4		BPSK	160 ms	13 dB
5		Up-chirp	40 ms	8 dB
6		BPSK	10 ms	0 dB
7	Time Domain Weighting	Combined	10 ms	2 dB

Table 3.5: Order of pulse audition in perceptual evaluation, and gain used to set all the pulses with the same SPL peak value.

3.7 Chapter Summary

This chapter focuses on non-invasive acoustic ranging. It starts with an overview of the acoustics theory that is used throughout this thesis to model the indoor acoustic environment. The indoor acoustic environment model includes the sound propagation model, acoustic transducer model, noise model and room model. A study on the perceptual impact of sounds on humans is then undertaken, starting with the characterization of the human hearing system, namely frequency response and loudness perception, in order to propose a series of non-invasive pulses for use in the acoustic ranging process. The audio input bandwidth for several commercially available smartphones is then characterized and the pulse detection theory for the matched filter case is introduced. In the end of this chapter we addressed the non-invasive pulse design approaches, followed by the description of the procedure used in the evaluation with subjects.

CHAPTER 4

WSN Infrastructure

The proposed system uses an infrastructure of acoustic beacons which must be coordinated to enable high accuracy ToF measurements and thus obtain high accuracy position estimates. The coordination stage, as presented in Chapter 2 (Figure 2.13), is critical in the ranging process and it is directly related to the synchronization approach and to the medium access strategies (viz. TDMA, FDMA and CDMA).

Range-based systems normally present heavy timing constraints only met using specific synchronization protocols in order to keep all the intervenient nodes in sync, i.e. under the same universal clock. To keep the acoustic WSN infrastructure in sync, a node with acoustic transmission and synchronization capabilities is proposed.

The simpler way to introduce synchronization into a range-based acoustic positioning system is to add a dedicated RF module (normally in the 433 MHz band), as the systems presented in [33] and [35]. In this case data communication is supported by the RF transceiver that is native to the WSN beacon, and a second RF module is included for synchronization purposes. As a result, hardware becomes more complex and costly and an extra RF channel is occupied.

The proposed architecture circumvents all of these problems by introducing an Automatic Time Synchronization and Syntonization (ATSS) protocol that takes advantage of the built-in RF transceiver for both communication and synchronization. Main advantages of the proposed approach include:

– Simple hardware design (fewer components);
– Cost reduction (only one radio);
– Reduces RF interference (fewer bands);
– Security increase (sync data can be encrypted).

4.1 Multiple Access

Multiple access is a term used in radio communications when the medium needs to be shared by multiple users in a common geographical area and specific time interval. In our case we need multiple ToF measurements for each beacon-target node pair. Multiple access can be performed using three main methods: Time Division Multiple Access (TDMA), Frequency Division Multiple Access (FDMA) or Code Division Multiple Access (CDMA). TDMA uses distinct time slots assigned to each distinct beacon and FDMA uses different frequency channels for each distinct beacon. CDMA uses both time and frequency diversity through different spreading codes, i.e. a different code is assigned to each distinct beacon.

In our application we are mainly focused on ToF measurements, and not on data communication. This requires that all the beacons in a specific room should be synchronized to the same reference clock to estimate ToF measurements.

Considering K beacons transmitting, each time slot can be split into three distinct intervals, namely:

$S^{(k)}$ — signal transmission interval;
$L^{(k)}$ — listening interval;
$G^{(k)}$ — guard time interval,
where the time slot duration for beacon k is given by $T_s^{(k)} = S^{(k)} + L^{(k)} + G^{(k)}$.

The signal transmission interval ($S^{(k)}$) is the time that the transmitter needs to send the acoustic pulse. The listening interval ($L^{(k)}$) is the time slice used by the target node to estimate the range measurement. For example, using a listening period of 30 ms results in distance measurements up to 10 m. The guard time interval ($G^{(k)}$) was added to reduce the impact of room reverberation.

Additionally, we have hard bandwidth limitations that are related to the audio I/O constraints in the hardware of conventional smartphones (maximum sampling frequency of 44.1 kHz) which creates an incompatibility with the use of FDMA due to the small bandwidth available. In this case we are restricted to an effective bandwidth between 4-6 kHz, which is near the Nyquist frequency limit. Due to this, we will design our WSN infrastructure having in mind the TDMA and CDMA approaches.

4.1.1 TDMA Approach

This method uses different time slots of a given length that are assigned to each beacon independently. The transmission is performed rotationally, with only one beacon transmitting at each distinct time slot. Figure 4.1 illustrates the time slot structure proposed for the coordination process when K beacons are used. For each beacon a specific time slot is reserved for signal transmission. In our approach, the slot synchronism for K beacons is achieved by means of a slightly longer guard time period in the last transmitted signal for a given TDMA structure, i.e. $G^{(K-1)} > G^{(k)}$

when $0 \leq k < K - 1$.

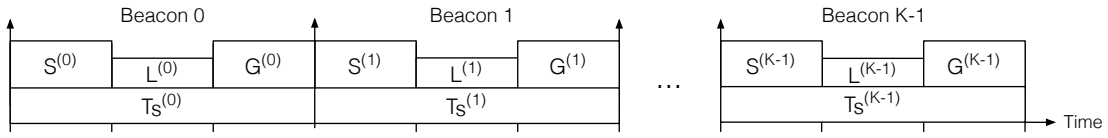


Figure 4.1: TDMA operation mode.

4.1.2 CDMA Approach

In a CDMA approach the frequency band is assigned to all beacons, which are able to transmit pulses at the same time, cf. Figure 4.2. The transmitted pulses are made distinguishable by assigning mutually uncorrelated code modulation patterns to the various beacons [89]. This method uses the same time slot of a given length assigned to all beacons to transmit pulses at the same time. The transmitted pulses are specially designed using quasi-orthogonal binary modulated sequences, such as Barker, Gold or Kasami sequences [89].

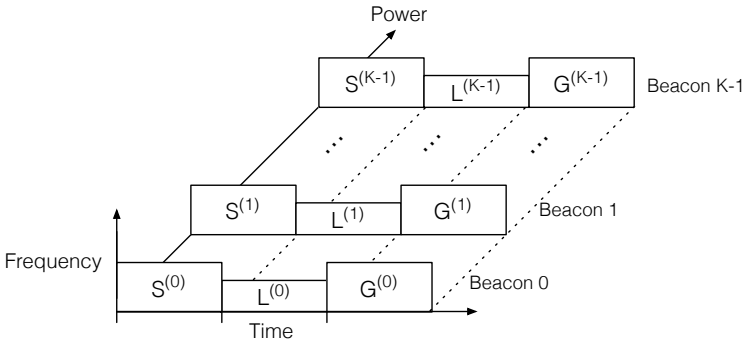


Figure 4.2: CDMA operation mode.

4.2 Synchronization in WSNs

One major problem faced in synchronization of a distributed WSN comes from the local oscillator in each node. Slight differences in the physical environment and in the hardware itself introduce significant changes in the local oscillators, thus resulting in clock drifts over time. The oscillator frequency deviation depends not only on the hardware but is also influenced by ambiental factors such as temperature, pressure or supply voltage. A reliable local oscillator has the ability to keep in sync over large periods of time (in the order of minutes). These large periods can be used for re-synchronization, because only one initial synchronization does not guarantee that devices keep eternally in sync. Moreover, Medium Access Control (MAC) delays and packet buffering result in non-deterministic packet delivery times, thus preventing WSN network protocols from being used to synchronize distributed events [90].

Note that a quasi-perfect clock can be obtained by keeping each node synchronized using the clock extracted from a GNSS receiver. However, this approach is expensive and useless in our case due to the difficulties that GNSS signals face in indoor environments [11, 12].

4.2.1 Time Synchronization Problem

Typically, WSN nodes are based in low-cost crystal oscillators with tolerances ranging from 10 ppm to 100 ppm, where ppm stands for part per million. This value can be interpreted as the number of microseconds deviation after one second. For example, according to the datasheet of a typical crystal-quartz oscillator commonly used in sensor networks, the frequency of a clock varies up to 40 ppm, which means that clocks of different nodes can diverge as much as 40 μs in one second [91].

4.2.2 Clock Definition

Theoretically, the clock of a sensor node can be described as $C(t) = t$, where t is the ideal or reference time. However, due to the imperfections of the clock oscillator, the clock function of the ith node can be described as,

$$C_i(t) = \phi + \omega t + \epsilon \tag{4.1}$$

where ϕ represents the clock offset (phase difference), ω is known as the clock skew (frequency difference) and ϵ represents random noise.

Figure 4.3 illustrates the clock drift over time for three distinct beacons, each one with distinct ϕ and ω. In this case, all nodes should synchronize to the reference clock (Beacon 1).

Assuming that the effect of random noise ϵ is negligible in 4.1, the clock relationship between the network nodes (Beacons 2 and Beacon 3) with a reference node (Beacon 1) can be represented by

$$C_2(t) = \phi^{(21)} + \omega^{(21)} C_1(t) \tag{4.2}$$

$$C_3(t) = \phi^{(31)} + \omega^{(31)} C_1(t) \tag{4.3}$$

where $\phi^{(ij)} = \phi^{(i)} - \phi^{(j)}$ is the relative clock offset and $\omega^{(ij)} = \omega^{(i)}/\omega^{(j)}$ is the relative clock skew between nodes i and j. When $\phi^{(i,j)} = 0$ and $\omega^{(i,j)} = 1$ the two clocks are perfectly synchronized, cf. Figure 4.3. Suppose there are K nodes in the network, then the global synchronization is achieved when $C_i(t) = C_j(t)$ for all $i, j = 1, \ldots, K$.

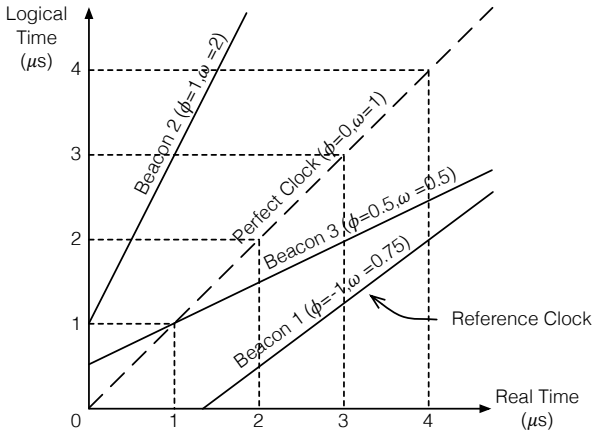


Figure 4.3: Illustration of the clock drift for three distinct acoustic beacons over time.

The clock skew ω is generally observed as a time-dependent random variable and two concepts are normally used in clock terminology to evaluate the clock drift over time, namely short-term and long-term stabilities. Short-term instability is mainly related to environmental factors, e.g. temperature variations, shock and supply voltage. On the other hand, long-term instability results mainly from oscillator aging [91].

WSN nodes typically use oscillators with all related parameters approximately constant for short-time intervals. Also, the total power of the noise is too small to be significant in short-time intervals, therefore the clock is assumed to be constant in the period of interest.

On the other hand, if the long-term stability is considered, the clock parameters are subject to changes due to environmental factors (temperature, atmospheric pressure, voltage changes, and hardware aging). In this case the relative clock offset keeps changing over time, which means that the network has to perform periodic time re-synchronization [91].

4.2.3 Synchronization Approaches in WSN

A time synchronization protocol should be designed based on requirements, such as energy efficiency, scalability, precision, security, reliability, and robustness to network dynamics. However, it is practically impossible to optimize a synchronization protocol to meet all these requirements simultaneously. This results in several protocol design approaches that can be divided into these three main categories [91]:

1. Master-slave vs. Peer-to-peer
   - Master-slave: first a tree-like network hierarchy is arranged, and upon the completion of this arrangement only the connected nodes in the hierarchy synchronize with each other.
   - Peer-to-peer: any pair of nodes in the network can synchronize with each other.

2. Clock correcting vs. Untethered clock

   – Clock correcting: the clock function in memory is updated after each run of the time synchronization process.

   – Untethered clock: every node maintains its own clock as it is, and keeps a time-translation table relating its clock to other nodes clocks. This way, instead of updating its clock constantly, each node translates the time information in the data packets coming from other nodes to its own clock by using the time-translation table.

   – Sender-receiver: one of two nodes, which are synchronizing with each other, sends a timestamp message while the other one receives it.

   – Receiver-receiver: a reference node transmits synchronization signals and two synchronizing nodes receive these signals and record the reception time instants (timestamps).

   – Receiver-only: a group of nodes can be simultaneously synchronized by listening to the message exchanges of a pair of nodes.

3. Pairwise synchronization vs. Network-wide synchronization

   – Pairwise synchronization: the protocols are primarily designed to synchronize two nodes, although they usually can be extended to handle synchronization of a group of nodes.

   – Network-wide synchronization: the protocols are primarily designed to synchronize a large number of nodes in the network.

The proposed synchronization approach follows the application requirements presented in chapter 1 and uses a master–slave topology with clock correcting performed by each node using a pairwise synchronization extended to a group of nodes in a two-level hierarchical spanning tree network (cf. Figure 4.4). In this topology, the clock reference is given by the main router node.

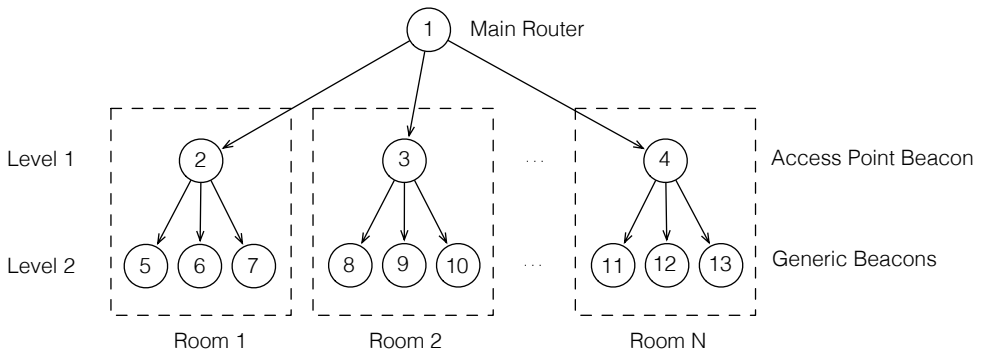


Figure 4.4: Clock propagation using a two-level hierarchical spanning tree network.

4.3 Automatic Time Synchronization Protocol

The IEEE 1588 standard [92] was developed for distributed networks requiring a better precision than that offered by the Network Time Protocol (NTP) when a GNSS module is not available, due to indoor use or cost constraints. With this protocol, it is possible to create a distributed network using a single GNSS module to provide a time reference for the network while maintaining accuracies with orders of magnitude of nanoseconds, cf. [93, 94].

The proposed ATSS protocol presents an internal architecture that relies in two main blocks, cf. Figure 4.5. The SimpliciTI[1] protocol enables data connectivity between beacons allowing the creation of star networks. Some changes have been made to allow recording the time of transmission and reception of messages as close as possible to its occurrence. The Time Sync block is responsible for managing the exchange of synchronization messages.

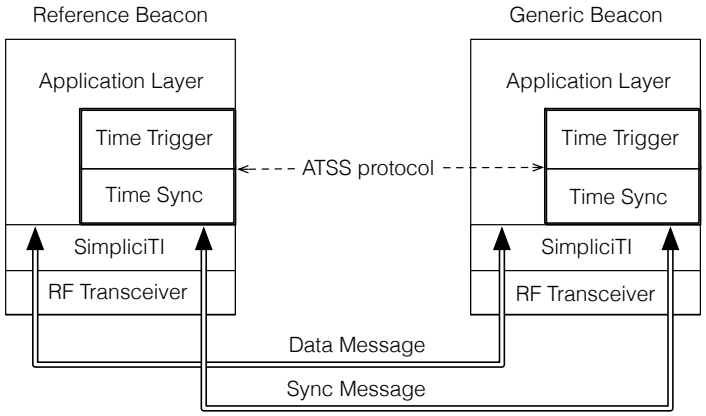


Figure 4.5: Beacon communication layers for synchronization and data exchange with ATSS protocol included.

This way the process of clock synchronization is performed transparently to the rest of the application. The Time Trigger block enables the creation of coordinated events between units with a resolution of 1 ms, but aligned in phase with the remaining units and it is also responsible for the adjustment of the syntonization period. The application runs at a higher level and has access to the communication channel to exchange messages, generate timestamps for detected events and schedule events to be triggered in coordination with other units.

[1] SimpliciTI™ [95] is a low-power RF protocol developed by Texas Instruments to be used in simple and small RF networks. It is an open-source software that was designed for easy implementation and deployment on several TI RF platforms such as the MSP430 MCUs.

4.3.1 Simplified Implementation of the IEEE 1588 standard

The IEEE 1588 standard specifies the type of message using a message ID byte with only five defined types [92]: Sync (ID 0), Delay Request (ID 1), Follow Up (ID 2), Delay Response(ID 3) and Management (ID 4), being the ID's between 5 and 255 reserved for future implementations. Due to physical limitations of the WSN nodes, a simple version comprised by the set of the first four previous defined messages (IDs 0–3) was implemented, cf. Figure 4.6. To achieve the best possible results, timestamps must be generated in hardware or as close as possible to the hardware and the transmitted messages composed by the minimum information necessary to proceed with the time synchronization.

The overall synchronization process occurs as follows:

- Sync [ID 0]: The synchronization process starts with the transmission of a broadcast "Sync" message containing a timestamp estimate of the reference beacon, which all remaining beacons use to register time t_2.
- Follow Up [ID 2]: Simultaneously, the reference beacon registers time t_1 after sending the Sync message and sends its value in the "Follow Up" message.
- Delay Req(uest) [ID 1]: Later, the generic beacon sends the "Delay Req" message and registers time t_3.
- Delay Resp(onse) [ID 3]: When the "Delay Req" message is received, t_4 is registered and transmitted back in the "Delay Resp" message.

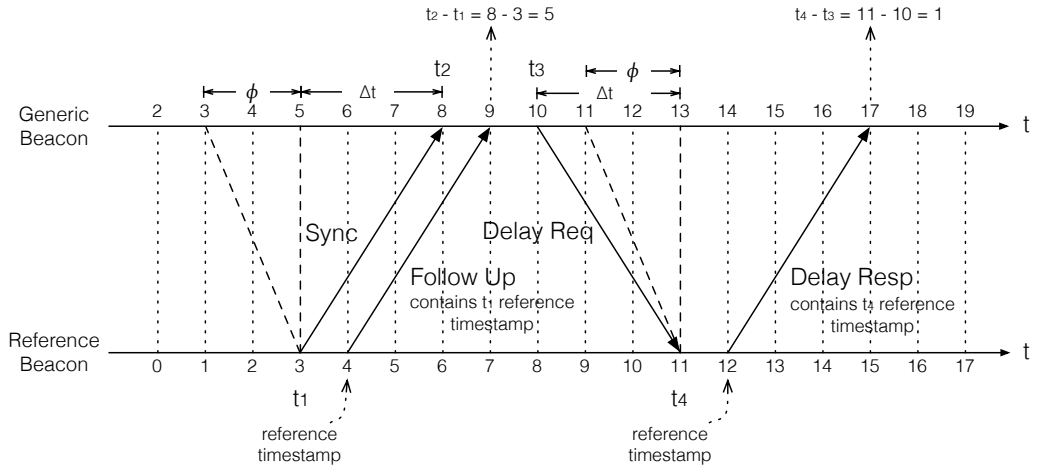


Figure 4.6: ATSS protocol message exchange in each synchronization point.

Based on "Sync" and "Follow Up" messages and in its own clock, the generic beacon computes the time difference between its own clock and the reference clock. The path delay propagation time is determined in a second transmission process between the beacons using the "Delay Req"

and "Delay Resp" messages. The generic beacon can then correct its clock and adapt it to the current path delay propagation time. Assuming a symmetric communication, the path delay can be obtained by $\Delta t = \frac{1}{2}[(t_2 - t_1) + (t_4 - t_3)]$, and the offset value ($\phi$) between the clock of the reference and the synchronized beacon is obtained by $\phi = (t_2 - t_1) - \Delta t$.

4.3.2 Clock Syntonization

The clock syntonization is performed by conducting small periodic corrections as shown in Figure 4.7. In this way the absolute value of the offset error is considerably reduced. The syntonization correction value (c_v) used depends essentially on the level of accuracy required by the application. The smaller this value, the greater the number of adjustments needed to perform between synchronization points. The number of necessary corrections (N_c) is related with the offset error after syntonization by $\phi_s = \frac{\phi}{N_c}$, cf. Figure 4.7.

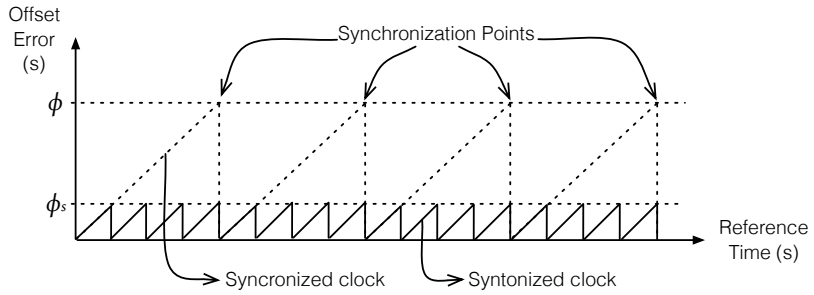


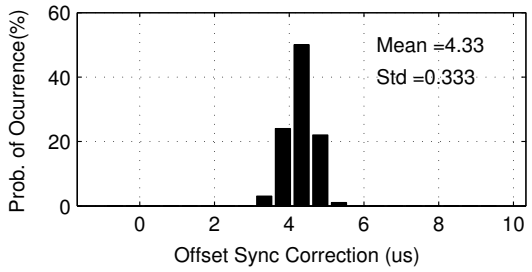
Figure 4.7: Improvement of clock offset error when using the ATSS protocol.

4.3.3 Protocol Evaluation

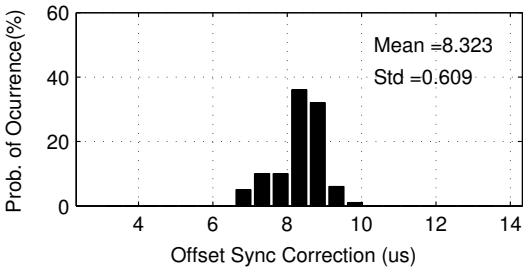
The presented protocol was evaluated using two beacons configured with a syntonization correction value of 1 μs and both connected to a pulse generator. For each transmitted pulse from the generator both units recorded the instant that the pulse arrived as a timestamp followed by its communication to a PC through a USART/USB connection. Two thousand measurements were taken to evaluate the protocol using the synchronization periods of: 1 s, 5 s, 10 s, 30 s and 60 s.

Figure 4.8 shows the offset correction obtained when syntonization is not used, for each synchronization period. In Figure 4.8g it is possible to observe the absolute offset correction value that should be performed to keep the nodes in sync. A linear regression of the acquired data shows the cumulative nature of the clock drift, diverging at a rate of approximately 3.9 μs/s. Figure 4.8h depicts the relative offset sync correction value for all synchronization periods tested.

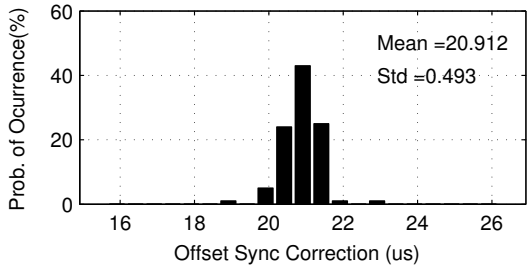
Figure 4.9 presents the evaluation of the proposed ATSS protocol for all synchronization periods tested. From the statistical data obtained it is possible to observe an absolute mean error always below 0.86 μs and a standard deviation always below 1.25 μs, which is more than sufficient for our application case.



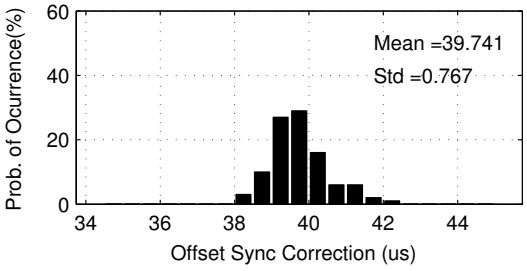
(a) Sync Period of 1 s.



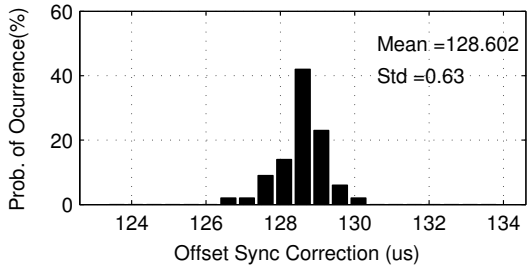
(b) Sync Period of 2 s.



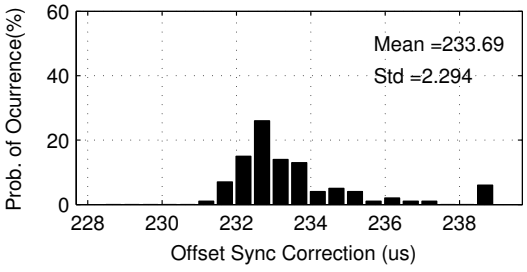
(c) Sync Period of 5 s.



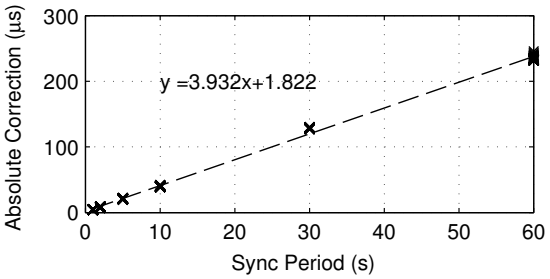
(d) Sync Period of 10 s.



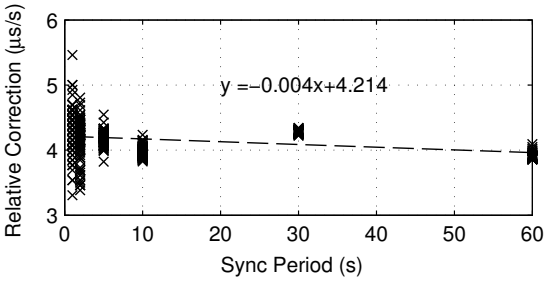
(e) Sync Period of 30 s.



(f) Sync Period of 60 s.

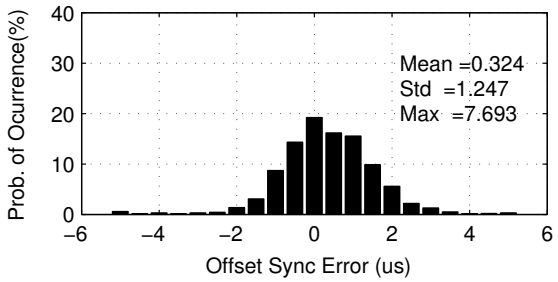


(g) Absolute offset sync correction for different synchronization periods.

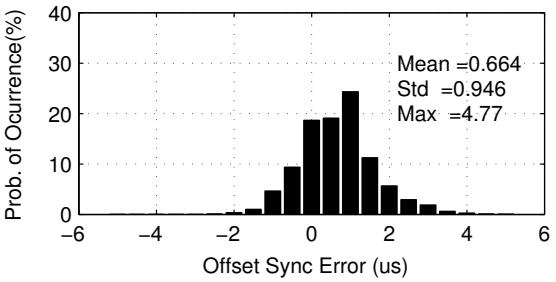


(h) Relative offset sync correction (μs/s) for different synchronization periods.

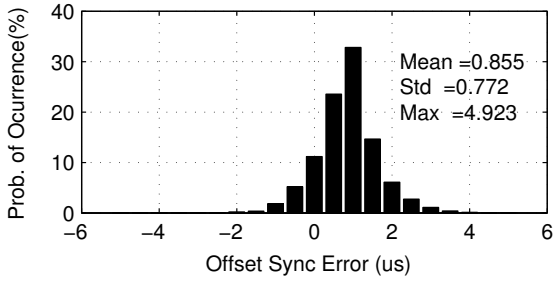
Figure 4.8: Offset sync correction obtained between two beacons for different synchronization periods without syntonization.



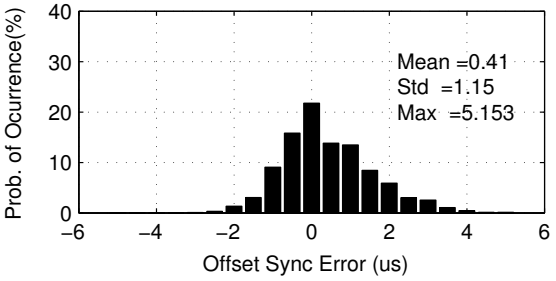
(a) Sync Period of 1 s.



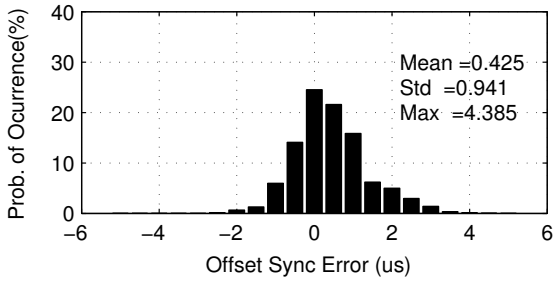
(b) Sync Period of 2 s.



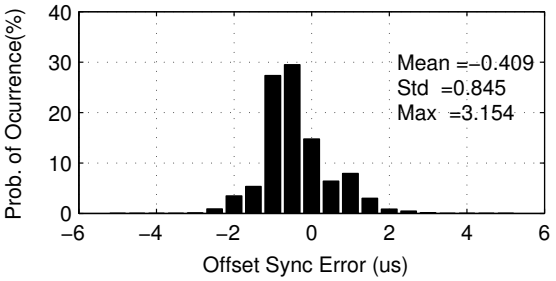
(c) Sync Period of 5 s.



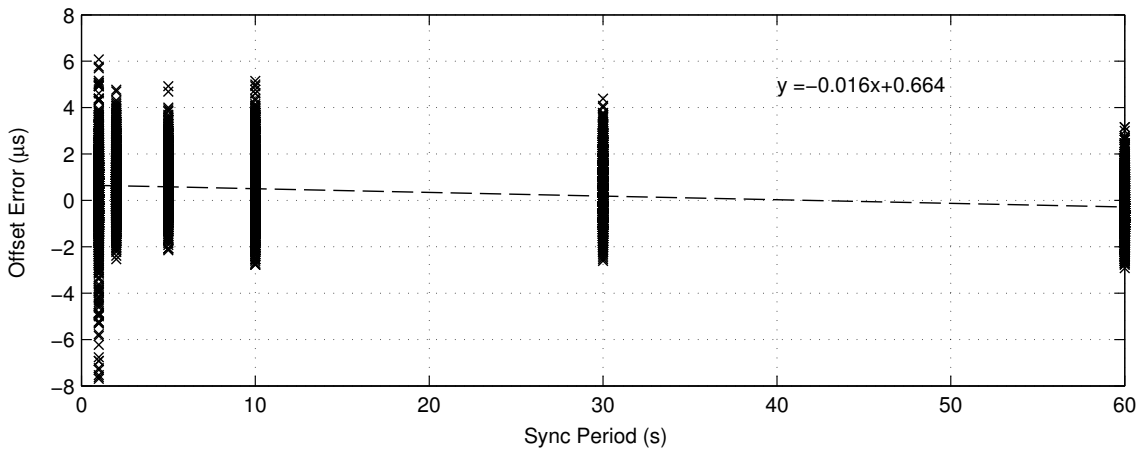
(d) Sync Period of 10 s.



(e) Sync Period of 30 s.



(f) Sync Period of 60 s.



(g) ATSS clock drift for different synchronization periods.

Figure 4.9: ATSS Protocol evaluation using two beacons for different synchronization periods.

4.4 Acoustic Beacon Design

The acoustic beacon is the atomic element of the WSN infrastructure proposed in this chapter, and plays a principal role in the ranging process. As specified early in this document, the system was designed to work in downlink mode, i.e. the beacons will work as transmitters and each target node will be able to obtain its own position estimate (self-positioning). A microphone was added to the beacon to enable mutual location, a foreseen advancement which simplifies and automates the system setup in a new room. In the beacon design process, the following requirements were followed:

- WSN-based: take advantage of the built-in WSN RF module for both communications and synchronization.
- Quasi-Omnidireccional and Broadband Transducers: This requirement enables the use of low density networks (with at least 3 acoustic beacons per room) with simple hardware requirements, such that a beacon node only needs a transmitter and a mobile node only needs a receiver.
- High Compatibility: the beacon should be compatible with different types of mobile nodes, namely smartphones, tablets and laptops (no special hardware is needed) or dedicated mobile nodes.
- Low Power Design: low power design techniques should be considered to prolong the beacon battery lifetime.
- Low Cost: Comercial-off-the-shelf (COTS) components must be used in the design process.
- Remote configurable: Each independent acoustic beacon should be controlled remotely through a common internet connection.

In the following sections the transducer characterization and the beacon architecture will be presented followed by its experimental validation.

4.4.1 Transducer Characterization

Transducers are critical devices in acoustic ranging systems due to several important factors, such as low-cost, commercial availability, directivity and power efficiency or sensitivity in the case of a transmitter or a receiver, respectively. We have opted for the COTS low-cost Kemo L10 piezo-tweeter and the Panasonic WM61-A electret microphone.

Piezo-tweeters are low-cost transducers commonly used in ultrasonic animal repeller systems and may easily be found at local stores. We measured the frequency response of several unbranded and commercially available piezo-tweeters before choosing one. In Figure 4.10 is presented the

frequency response versus the sound pressure level per Volt (SPL/V) of the selected piezo-tweeter, the Kemo L10. This graph was obtained when the transmitter and receiver were at 1 m of distance, considering the zero-degree case of its radiation pattern. We have used a Brüel & Kjær 4954A reference microphone for measurement and calibration, cf. Appendix A. Looking at the piezo-tweeter frequency response it is possible to output an average acoustic power of almost 80 dB(SPL)/V in the band of interest (15–30 kHz).

Figure 4.11 shows the sensitivity of the selected WM61-A microphone over frequency. This response was obtained after calibration with the same Brüel & Kjær 4954A reference microphone. Looking to Figure 4.11 in more detail, one can observe that the sensitivity is almost constant ($S \approx -35$ dB) in the interest band (15–30 kHz).

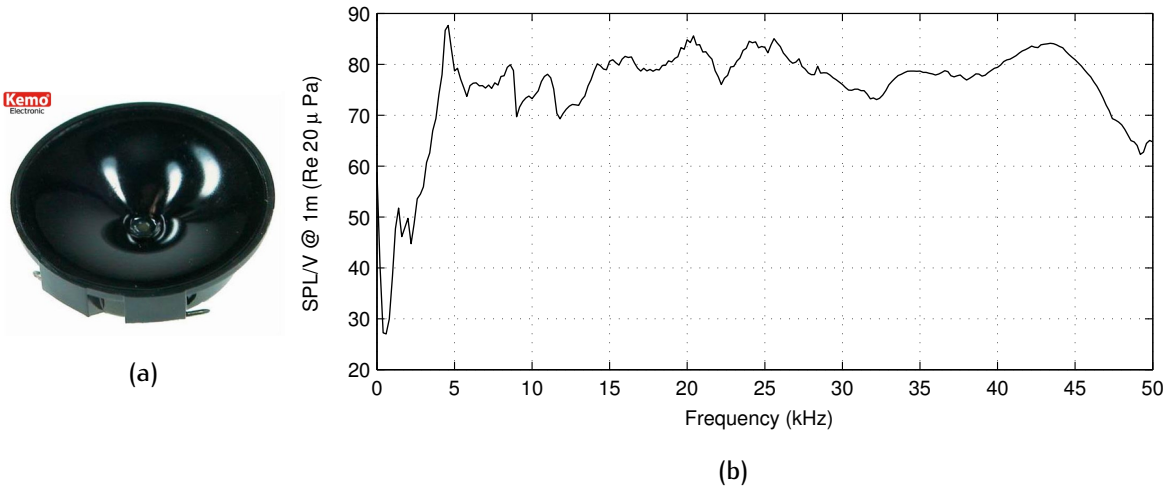


Figure 4.10: (a) Kemo L10 piezo-tweeter. (b) Measured frequency response.

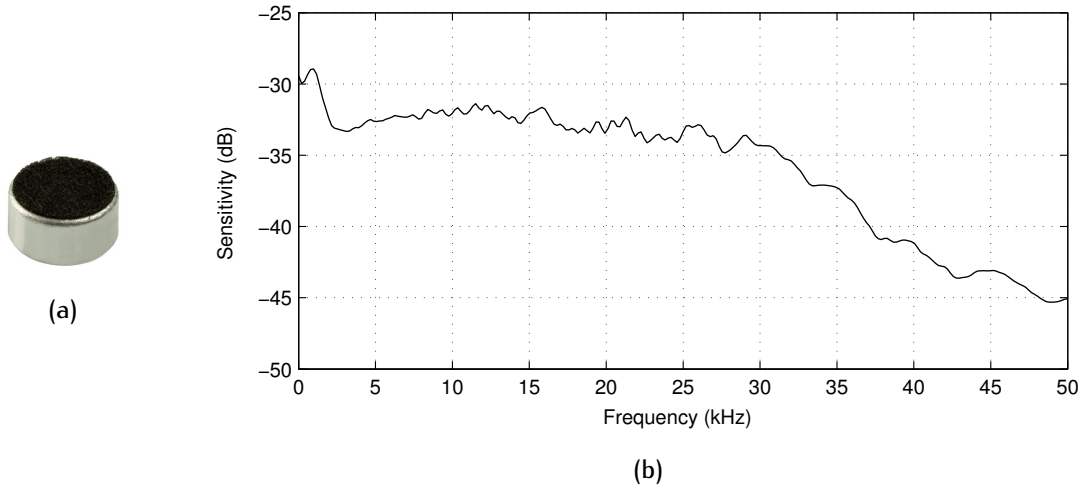
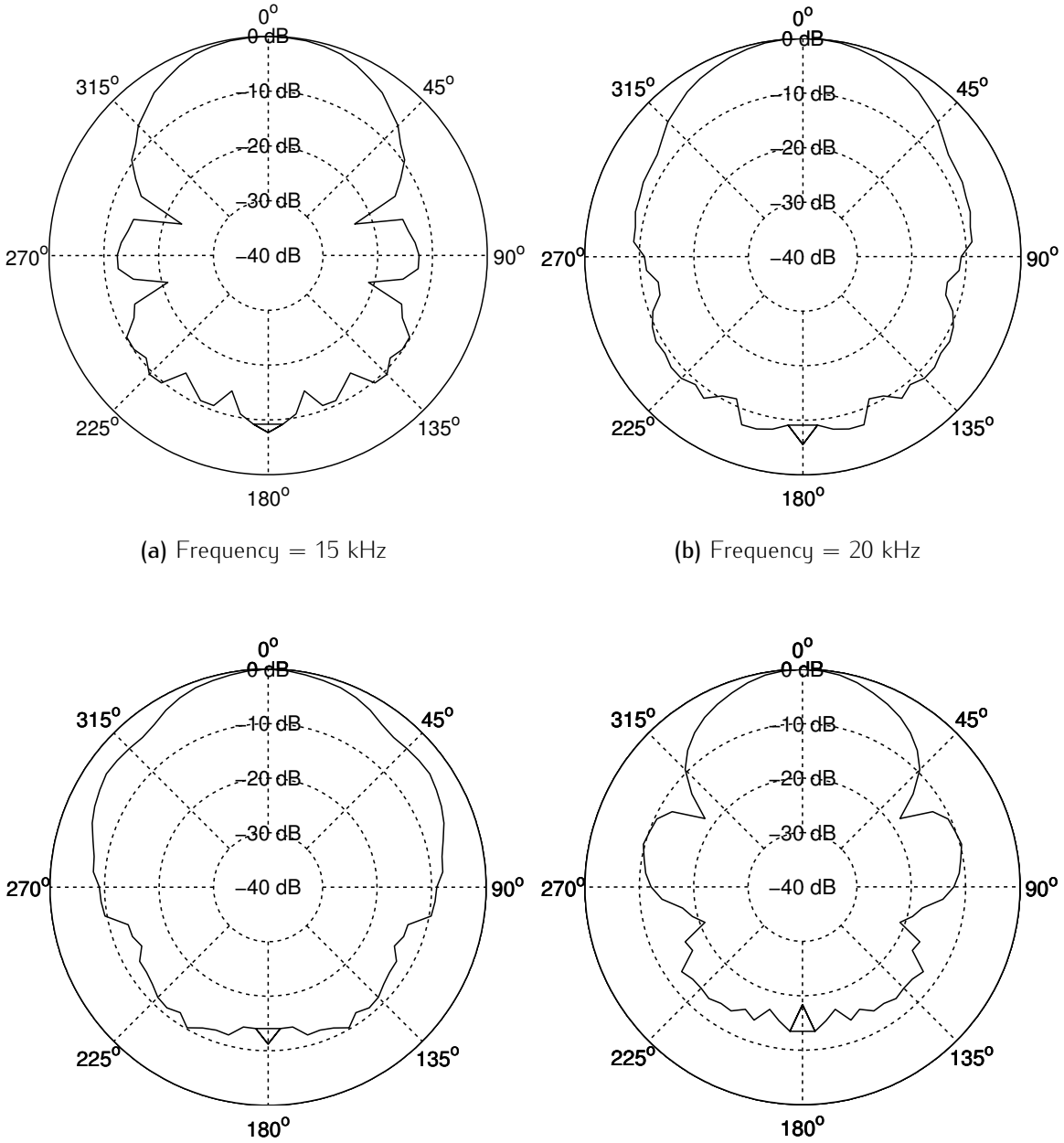


Figure 4.11: (a) Panasonic WM61-A electret microphone. (b) Measured frequency response.

By considering a symmetric radiation pattern for the Kemo L10 piezo-tweeter, several measurements were taken in steps of 10° for the interval 0° and 180° which resulted in the radiation pattern presented in Figure 4.12 at four discrete frequencies in the band of interest (f_1 = 15 kHz, f_2 = 20 kHz, f_3 = 25 kHz and f_4 = 30 kHz).



(a) Frequency = 15 kHz

(b) Frequency = 20 kHz

(c) Frequency = 25 kHz

(d) Frequency = 30 kHz

Figure 4.12: Kemo L10 piezo-tweeter beam pattern, measured for different frequencies.

4.4.2 Beacon Architecture

Figure 4.13 presents the overall architecture of the proposed acoustic beacon. Two main modules are introduced: the communications module and the acoustic module. The communications module deals with the communications and synchronization tasks. It implements the RF protocol for exchanging messages and, on top of it, the previously introduced ATSS protocol. The acoustic module is responsible for acoustic signal generation and transmission, based on the received remote configuration messages and external interrupts used as acoustic transmission triggers.

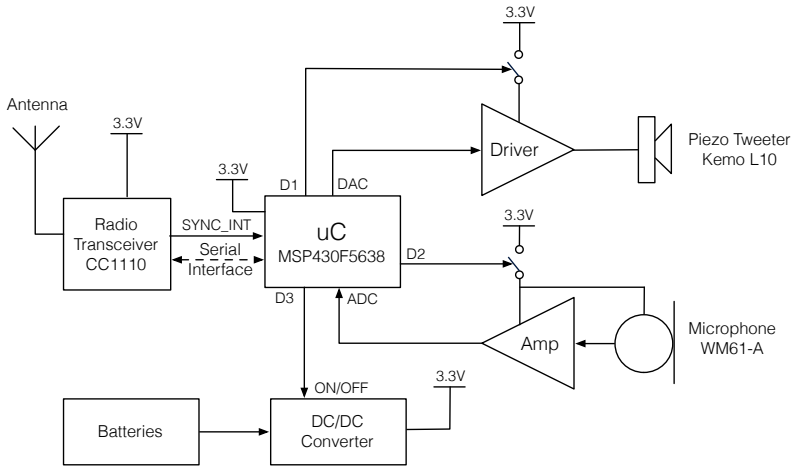


Figure 4.13: Overall acoustic beacon architecture

Both modules are interconnected by a serial communication port (used for data exchange) and a digital signal (SYNC_INT) working as an external interrupt to trigger the acoustic transmission. The communications module is based on a System-on-Chip (SoC), the CC1110 from TI, a low-power microcontroller with a built-in low-power RF transceiver for network communications. This SoC is compatible with different network protocol stacks and is widely used with the SimpliciTI (TI open source stack) and other IEEE 802.15.4 based stacks. We have opted for the SimpliciTI protocol because it allows to use the 433 MHz band for communication. This way it is possible to communicate over longer distances with reduced building interferences, due to the longer wavelength of the RF signals.

The acoustic module integrates an ultra-low power (but powerful) microcontroller (MSP430F5638 from TI) with built-in A/D and D/A modules that are used for acoustic signal acquisition and generation, respectively. An op-amp with high output current is used to drive the piezo-tweeter during the acoustic pulse transmission interval. Low-power features have been used in the design of the acoustic board. All external modules include a shutdown functionality to reduce the energy consumption of the overall module, when not being used. The circuit schematic and the PCB design are presented in Appendix E.

Low-power programming techniques were used to increase the battery life of the system. Figure 4.14 presents the flowchart that describes the kernel running in each acoustic beacon. The

acoustic module wakes up whenever an internal (TIMER_INT) or external interrupt (USART_RX_-INT or SYNC_INT) occurs as the result of a Wake-on-Interrupt (WoI) implemented feature. After acoustic transmission the hardware used during transmission is turned OFF and the microcontroller is set to a low power consumption profile, i.e. SLEEP mode.

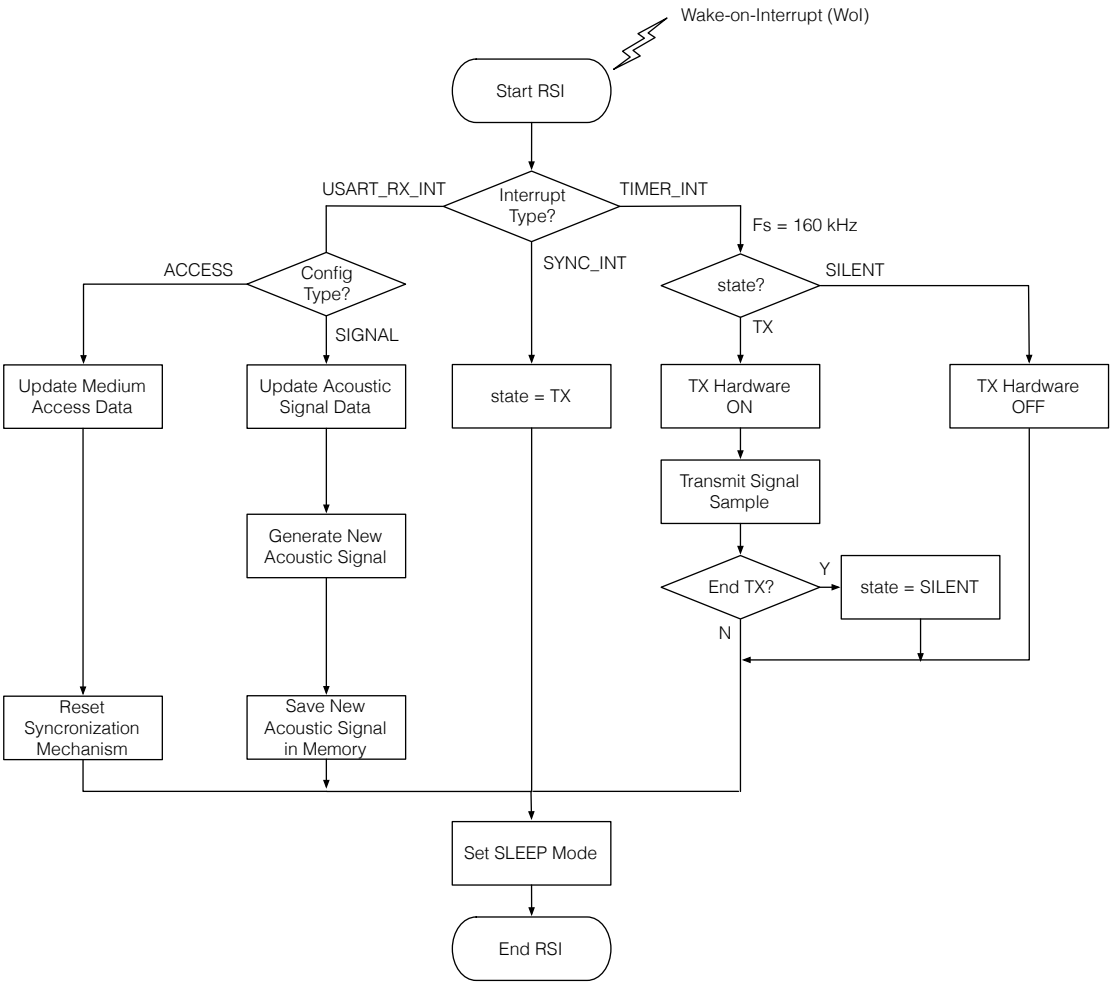


Figure 4.14: Acoustic module interrupt service routine flowchart.

The acoustic beacon was designed to include remote configuration through the wireless network. This feature enables the possibility to customize each acoustic node using on-the-fly acoustic signal generation. This way, acoustic signal parameters, such as signal type, bandwidth, duration or envelope can be assigned individually, and multiple access technique features (TDMA and CDMA) can rapidly be changed, to adapt the transmitted acoustic signals and the multiple access technique to a particular room with specific physical constraints.

The signal generation process is performed at each beacon on power up. Equation 4.4 was used to generate the chirp-based signals at the beacon,

$$x(t) = \sin\left(2\pi f_i(t)t\right) \tag{4.4}$$

where $f_i(t)$ defines the chirp type as described in Table 4.1, f_0 is the chirp starting frequency (at $t = 0$) and β is the assumed frequency increase rate.

Chirp Type	Parameters
(1) Linear	$f_i(t) = f_0 + \beta t$ where $\beta = (f_1 - f_0)/t_1$
(2) Quadratic	$f_i(t) = f_0 + \beta t^2$ where $\beta = (f_1 - f_0)/t_1^2$
(3) Logarithmic	$f_i(t) = f_0 \times \beta^t$ where $\beta = [\frac{f_1}{f_0}]^{\frac{1}{t_1}}$

Table 4.1: Chirp types that can be generated by the acoustic beacon.

A BPSK modulator was also added to the acoustic beacon. The idea is to turn the acoustic beacon compatible with the use of BPSK modulated quasi-orthogonal codes, such as Barker, Kasami or Gold sequences. The parameters needed to generate a BPSK sequence in the beacon are: the binary code sequence, the carrier frequency f_c, the sampling frequency f_s, and the number of periods required to transmit a bit N_c.

For easy implementation of the BPSK modulator at the beacon, we have selected a bit rate that is sub-multiple of the carrier frequency f_c, and a sampling frequency f_s that is multiple of f_c. With this, we have obtained a wave shape that is symmetrical at each phase transition with N_c periods per bit, cf. Figure 4.15.

The BPSK modulator realizes the modulation of each bit of the code sequence by transmiting N_c cycles of the carrier when a one is being modulated, or N_c cycles of the carrier with a phase inversion of π when a zero is being modulated

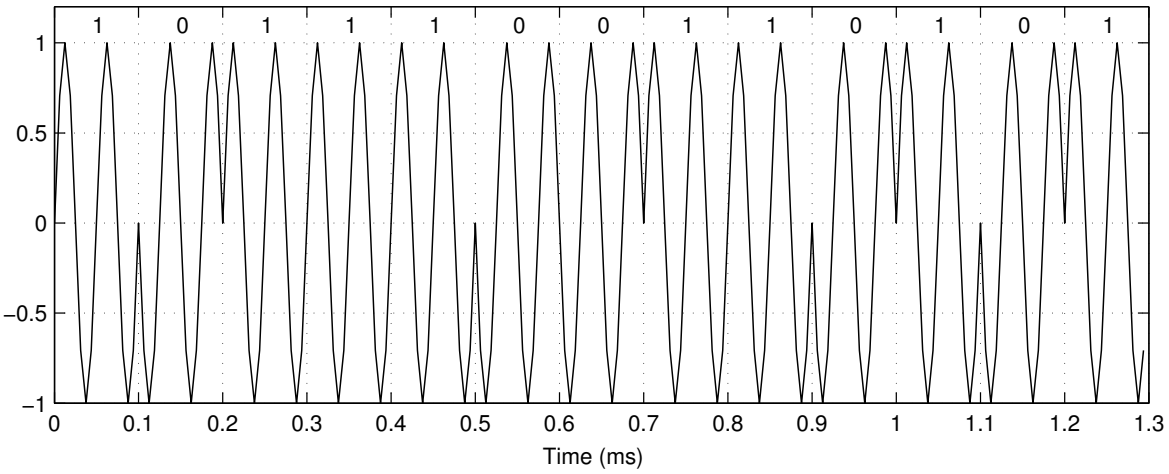


Figure 4.15: Example of a Barker code with length 13 modulated using BPSK, with f_c=20 kHz, N_c=2 and f_s = 160 kHz.

4.4.3 Prototypes

Figure 4.16 shows the developed prototypes, two acoustic beacons and a PC gateway for remote configuration. In the figure it is possible to observe the two main modules used in the beacon prototype (the communications module and the acoustic module) and the transducers used (the piezo-tweeter Kemo L10 and the microphone WM61-A).

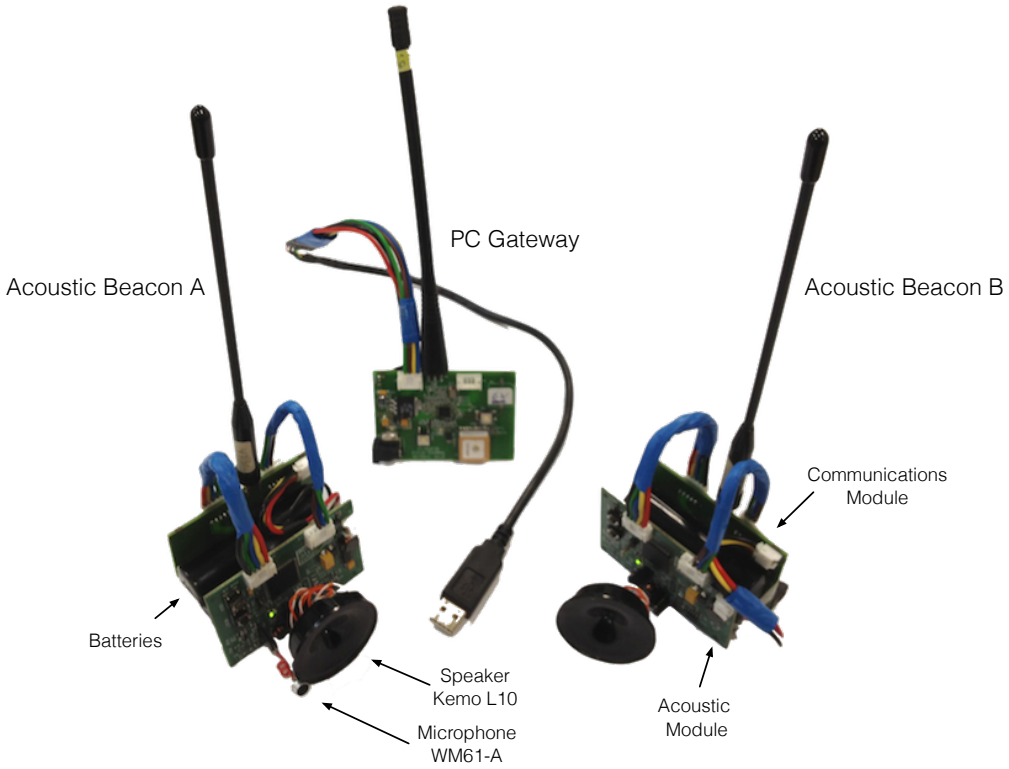


Figure 4.16: Acoustic Beacon Prototype.

To evaluate the battery autonomy of the beacon we measured the current consumption when the beacon is running the ATSS protocol with a sync period of 30 s and transmitting ten acoustic chirps per second (bandwidth 20–45 kHz and duration 10 ms) at a sample frequency of 160 kHz. Working in this mode the average consumption of the overall prototype was approximately 4 mA. Using AA Lithium batteries with typical capacities of 3400 mAh, the prototype presents an autonomy of approximately 36 consecutive days.

4.4.4 Experimental Validation

Two acoustic beacons were used for experimental ranging validation. A first beacon was configured to work as a transmitter sending an acoustic pulse every 500 ms and a second beacon was configured to work as a clock reference on the acquisition side to estimate the ToA of the acoustic pulse, cf. Figure 4.17.

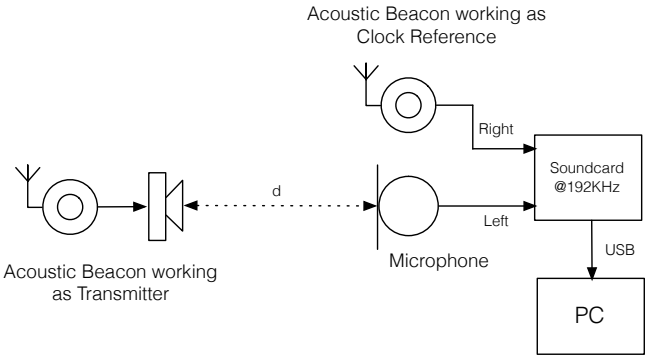


Figure 4.17: Experimental validation using two acoustic beacons and a PC for data acquisition.

A set of four experiments were performed using different types of acoustic pulses to measure the distance between two beacons. In all cases, the transmitted pulses were modulated in amplitude with a tapered cosine window (i.e. a Tukey window). A Tukey window is a rectangular window with the first and last $r/2$ fractions of samples equal to parts of a cosine, were r is a real number between 0 and 1; for all the experiments we have set $r = 0.3$. All transmitted pulses in all experiments were generated by an acoustic beacon with the duration of 10 ms and a sampling frequency of 160 kHz. At the receiver, a sampling frequency of 192 kHz was used for signal acquisition. The received acoustic pulse is detected on the left channel of the soundcard after correlation with the reference signal (local version of the transmitted signal), using a time domain matched filter to obtain the ToA estimate, by searching the maximum amplitude value of the correlated signal in a slot of 10 ms, after the sync pulse of the reference beacon appears at the right channel of the soundcard. The ATSS sync protocol was configured with a sync period of 5 s and a syntonization period of 1 μs.

The ranging experiments performed used the following pulses:

- Experiment A: A linear up-chirp with $f_0 = 18$ kHz and $f_1 = 24$ kHz (B = 6 kHz);
- Experiment B: A linear up-chirp with $f_0 = 20$ kHz and $f_1 = 40$ kHz (B = 20 kHz);
- Experiment C: A quadratic Chirp with $f_0 = 20$ kHz and $f_1 = 40$ kHz (B = 20 kHz);
- Experiment D: A BPSK 13-bit Barker Code with B = 20 kHz, f_c=30 kHz. The code used was the 13 bits Barker code given by the sequence $\{1, 1, 1, 1, 1, 0, 0, 1, 1, 0, 1, 0, 1\}$.

In each experiment a set of one thousand distance measurements were taken for each of the following distances d=1, 2, 4 and 8 m, with the transmitter and receiver in LoS at zero-degrees in both radiation patterns. The transmitter and the receiver were placed at 1.5 m above the floor to avoid interference of the floor echo. The results obtained for these experiments are presented in Table 4.2.

	Exp. A	Exp. B	Exp. C	Exp. D
d_r (m)	$\hat{d}_e \pm \sigma_e$ (m)	$\hat{d}_e \pm \sigma_e$ (m)	$\hat{d}_e \pm \sigma_e$ (m)	$\hat{d}_e \pm \sigma_e$ (m)
1	1.002 ± 0.008	1.003 ± 0.005	1.002 ± 0.004	1.003 ± 0.006
2	2.006 ± 0.008	2.007 ± 0.005	2.006 ± 0.003	2.007 ± 0.008
4	4.013 ± 0.009	4.011 ± 0.006	4.013 ± 0.003	4.011 ± 0.006
8	8.025 ± 0.008	8.023 ± 0.005	8.025 ± 0.004	8.021 ± 0.041

Table 4.2: Ranging results obtained with two beacons. d_r represents the real distance and $\hat{d}_e \pm \sigma_e$ are the average estimated distances and respective standard deviation.

Figure 4.18 presents the absolute error of the results presented in Table 4.2. From the figure it is possible to observe an absolute error always below 4 cm and typical standard deviations below 1 cm.

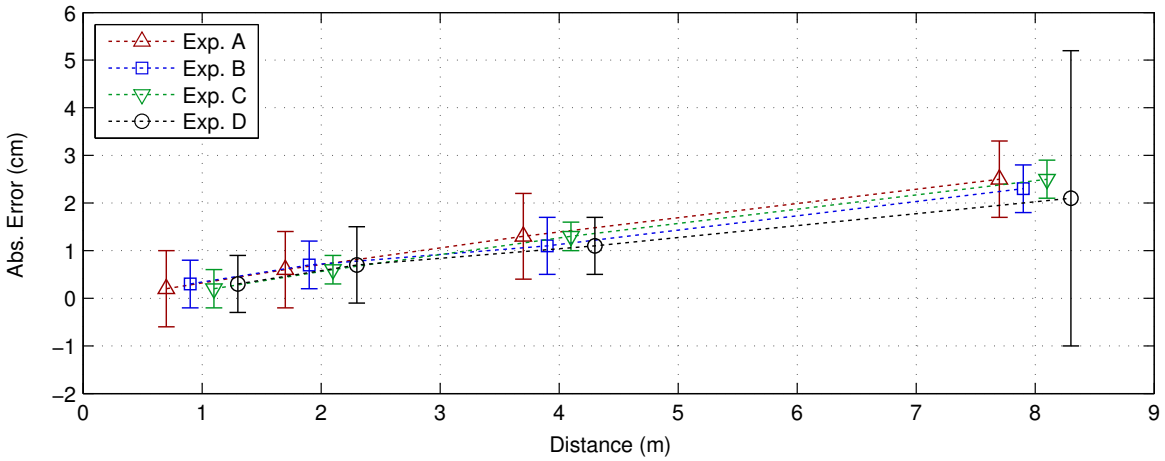


Figure 4.18: Absolute error for the experimental data in Table 4.2.

4.5 Chapter Summary

In this chapter we have proposed the design of an infrastructure of synchronized acoustic beacons to enable ToF measurements and thus obtain high accuracy position estimates. First, the coordination methods are introduced in relation to the medium access and then the proposed synchronization strategy is addressed. In the end the architecture of the acoustic beacon is introduced and its experimental validation is performed.

CHAPTER 5

Akkurate: Smartphone-based Head Positioning

The proposed system takes advantage of a WSN infrastructure and non-invasive audio pulses for low-cost ranging, enabling effective indoor positioning of conventional smartphones. The system uses an improved MT-TDoA estimation approach, exploiting the periodicity of the beacon signals' and performing Non-Line-of-Sight (NLoS) mitigation, to obtain accurate ranging measurements. In this way, mobile devices do not need to be synchronized with the positioning infrastructure, and a position estimate is obtained through a hyperbolic positioning approach.

The WSN infrastructure can be used in a TDMA or a CDMA schema. In both schemas the beacons transmit synchronous and periodic frequency modulated pulses, i.e. chirp pulses. The usage of chirp pulses overcomes most problems of pure sine tones, such as: poor time resolution, low environment noise immunity, short range, and low robustness to the Doppler effect. The probability of detection of a transmitted chirp is directly related with the SNR rather than the exact waveform of the received signal [87].

This chapter is organized as follows. First, the architecture of the proposed system is introduced, followed by a detailed description of the measurement and positioning approaches proposed. In the end of the chapter an experimental validation is performed using an iPhone 4s to evaluate the proposed system in real applications. Four acoustic beacons were used in a room of approximately 100 m^2. Two experiments were taken to evaluate the system performance. The first experiment was held to obtain a quantitative evaluation of the overall system, by measuring the position estimated in a grid of points in a regular room, and a second experiment was taken to obtain a qualitative evaluation of the positioning system when a person equipped with a mobile device is in a moving trajectory. Very stable 2D position estimates were obtained (absolute mean standard deviation less than 2.2 cm when four acoustic beacons were used) and a position refresh rate of 350 ms was achieved.

5.1 System Design

In the following subsections we focus on the requirements imposed by the application type we want to satisfy. In addition, a set of assumptions are made to support our design decisions.

5.1.1 System Requirements

The proposed system has been developed having in mind applications that require increased accuracy, in the centimeter-level order, such as augmented reality, virtual reality, gaming, and audio guiding applications. To achieve these requirements we have defined the following criteria:

- Indoor Operation: This requirement limits the use of GPS systems, due to attenuation, multipath and interference that RF signals suffer when used indoors.
- Smartphone Compatibility: The system must be compatible with conventional smartphones. This restricts the selection of the sampling frequency of the acoustic signal to the smartphone hardware constraints. Commercially available smartphones allow a maximum sampling rate of 44.1 kHz therefore limiting the useful acoustic upper band to the Nyquist frequency, i.e. 22.05 kHz. In addition, to avoid synchronization with the beacon infrastructure, TDoA measurements must be used to perform hyperbolic positioning.
- Centimeter-level Head Tracking: To satisfy this criterion a range-based positioning system using ToF measurements is needed. Additionally, an infrastructure of acoustic beacons with known positions must be used, to circumvent the lack of accuracy of mutual positioning systems [9].
- High Doppler Tolerance: The system should work when mobile stations are moving with velocities up to 2 m/s. This criterion imposes the usage of Doppler resilient acoustic signals.
- Low-Cost Infrastructure: A low density WSN should be used with at least three beacons per room to obtain 2D position estimates. The system infrastructure must be based in low-cost and comercial-off-the-shelf components.
- System Scalability: The system must work in downlink mode (GPS-like). Furthermore, the WSN infrastructure should use a star topology per room and rooms must be interconnected through a main router.

The WSN infrastructure introduced in Chapter 4 enables smartphone ranging only using acoustic signals, since the beacons are synchronized using the proposed ATSS protocol. In this case, MT-TDoA can be used to perform hyperbolic positioning. With this approach, it is possible to extend the system to other devices, such as tablets, laptops, or any device that has a built-in microphone and processing capabilities.

5.1.2 Proposed Architecture

Figure 5.1 illustrates the overall architecture of the proposed positioning system. We followed a modular approach, based in a two-level hierarchical spanning tree WSN to guarantee scalability. In this way, multiple rooms with unique IDs can be added depending on the needs. Each room forms a star network centred around an access point beacon that connects to the remaining generic beacons in the room. A gateway node is used as a bridge to interconnect the system to a WiFi router. Each mobile device is able to compute its own position followed by its communication to a remote positioning server through a standard data connection (WiFi or 3G/4G). Information about the system configuration (e.g. beacon positions and multiple access characteristics), and the actual position of each mobile device is stored in a remote database that may be accessed with a browser through a common internet connection.

Additionally, an application server is used to handle data related to the iOS app prototype we have developed. The app is called Akkurate and was developed to support an audio augmented reality system that is capable of generating virtual binaural sound sources in real-time. The data related to the virtual sound sources configuration is accessible through a web API developed in PHP.

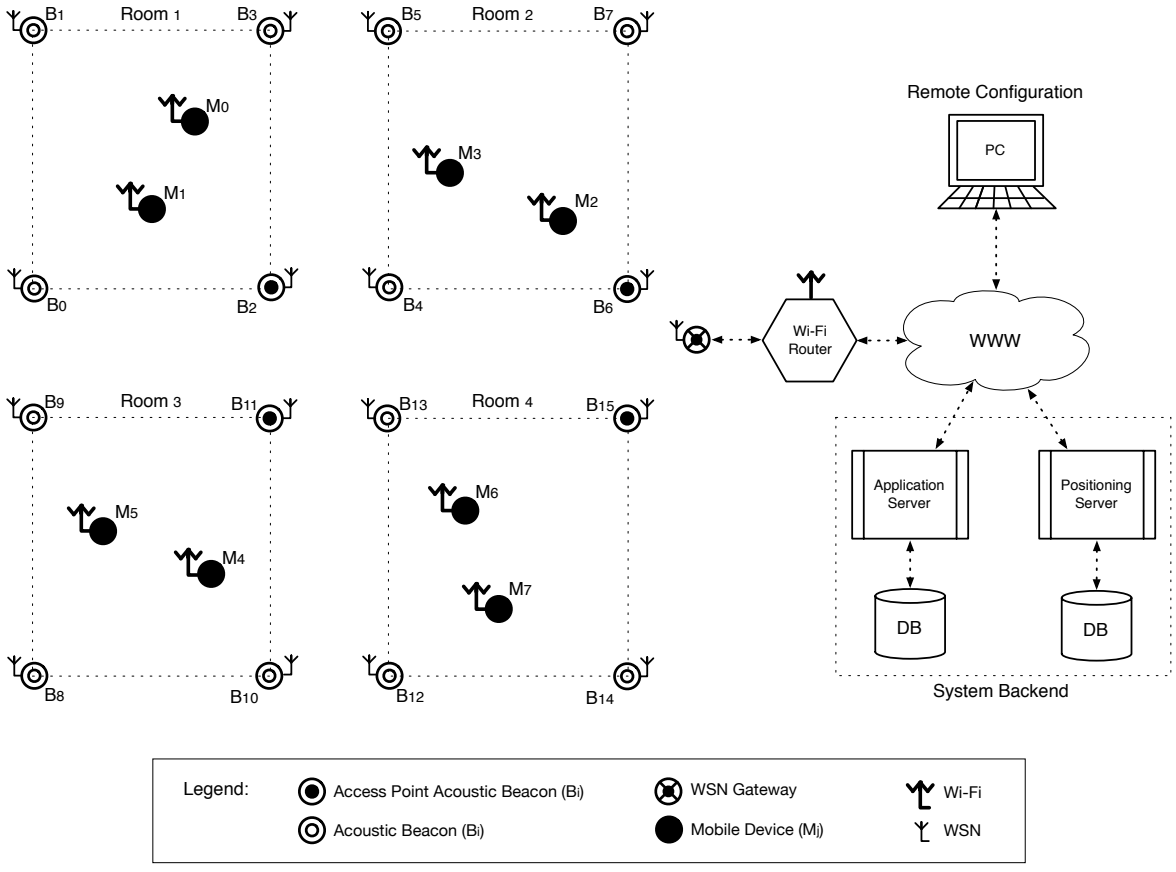


Figure 5.1: Overall system architecture.

5.2 Measurement Strategy

Distance measurement between network nodes will be based on MT-TDoA measurements. With this approach, synchronization is only mandatory between beacons because the clock bias can be canceled by taking the difference of the instants of arrival of the incoming pulses transmitted by distinct beacons. Thus, it enables low-cost ranging, because it is fully compatible with the built-in hardware of conventional smartphones.

5.2.1 MT-TDoA Ranging

MT-TDoA measurements enables the use of hyperbolic positioning. In the example illustrated on Figure 5.2, all beacons (B_i) transmit synchronized pulses at instant t_0. The target node will receive these pulses at different instants τ_i were i denotes the beacon index ($0 \leq i < K$ with K the number of beacons in use). Figure 5.3 presents a detailed timing diagram for the situation described in Figure 5.2.

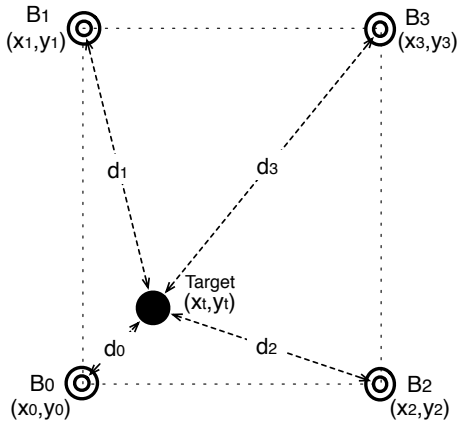


Figure 5.2: Positioning example using MT-TDoA measurements.

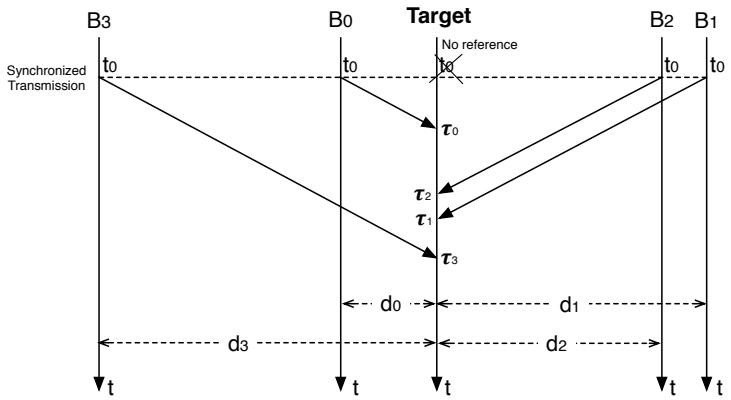


Figure 5.3: MT-TDoA approach, requiring no synchronization at the target device.

Assuming that beacon B_i is transmitting signal $s[n]$ from position (x_i, y_i), and the target node is at unknown position (x_t, y_t), the received signals at the target node can be modeled as a function in discrete-time n,

$$z_i[n] = A_i s(n - \tau_i) + w[n], \quad 0 \leq i < K \tag{5.1}$$

where A_i is the amplitude of the received signal, τ_i is the instant of arrival of the pulse at the receiver, and $w[n]$ is a noise function that can be modeled with AWGN.

In Figures 5.2 and 5.3, d_i represents the distance between the target node and the ith beacon. Given this, we can formulate the following basic squared distances d_i^2 relations:

$$d_i^2 \quad = \quad (x_i - x_t)^2 + (y_i - y_t)^2, \quad 0 \leq i < K \tag{5.2}$$

In the MT-TDoA approach, the received signals are compared pairwise, due to the lack of a clock reference at the target node. These pairwise measurements, i.e. MT-TDoA estimates, can be directly computed with $\mathrm{T}_{i,j} = \tau_i - \tau_j$. For the example presented in Figures 5.2 and 5.3, the relationships between the range difference $D_{i,j}$, the MT-TDoA estimates ($\mathrm{T}_{i,j}$), and the instant of arrival τ_i of the incoming pulses at the target node, are given by

$$D_{i,j} = d_i - d_j = v_{air}\mathrm{T}_{i,j} = v_{air}(\tau_i - \tau_j), \quad 0 \leq i < j < K \tag{5.3}$$

where $D_{i,j}$ is the distance difference between beacons B_i and B_j, d_i represents the euclidean distance between the target node and beacon B_i, v_{air} is the speed of sound in the air, $\mathrm{T}_{i,j}$ represents the time difference of arrival between the received signals (MT-TDoA estimates), τ_i the instant of arrival of $z_i[t]$, and (i, j) is an enumeration of all K pairs of beacons, where

$$M = \frac{K!}{2!(K-2)!} \tag{5.4}$$

For the example presented in Figure 5.2, we have $K = 4$, resulting in $M = 6$ distinct distance differences of arrival measurements,

$$D_{1,0} = d_1 - d_0 = v_{air}\mathrm{T}_{1,0} = v_{air}(\tau_1 - \tau_0) \tag{5.5}$$

$$D_{2,0} = d_2 - d_0 = v_{air}\mathrm{T}_{2,0} = v_{air}(\tau_2 - \tau_0) \tag{5.6}$$

$$D_{3,0} = d_3 - d_0 = v_{air}\mathrm{T}_{3,0} = v_{air}(\tau_3 - \tau_0) \tag{5.7}$$

$$D_{2,1} = d_2 - d_1 = v_{air}\mathrm{T}_{2,1} = v_{air}(\tau_2 - \tau_1) \tag{5.8}$$

$$D_{3,1} = d_3 - d_1 = v_{air}\mathrm{T}_{3,1} = v_{air}(\tau_3 - \tau_1) \tag{5.9}$$

$$D_{3,2} = d_3 - d_2 = v_{air}\mathrm{T}_{3,2} = v_{air}(\tau_3 - \tau_2) \tag{5.10}$$

Each previous $D_{i,j}$ value corresponds to a target node position (x_t, y_t) along a hyperbola. To solve for 2D we only need two of the above measurements to draw its two correspondent hyperbolas an thus solve for (x_t, y_t). To obtain 2D position estimates, a minimum of three beacons is required to uniquely determine the unknown position of the target node. If four beacons are used, it is possible to obtain three $D_{i,j}$ estimates, which gives the possibility of obtaining 3D position estimates. For simplification we focus on 2D, but the reader should be aware that all theory here introduced can easily be extended to 3D.

An important part of the MT-TDoA ranging process is the estimation of the instant of arrival τ_i of the incoming ith pulse at the target node. In the next subsections we will describe in detail the algorithms used to estimate τ_i and validate the MT-TDoA measurements.

5.2.2 Pulse Instant of Arrival Estimation

A matched filter is used to maximize the SNR before instant of arrival estimation, cf. Section 3.4.2. Figure 5.4 presents the flowchart of the algorithm used to estimate the pulse instant of arrival, to obtain the MT-TDoA estimates.

After matched filtering, a decimated energy estimator E_x based on the L^2-norm of the correlated signal R_{xs} is computed with the following equation,

$$E_x[i] = \sqrt{\sum_{n=0}^{L-1} |R_{xs}[n + i.L]|^2}, \ \forall i \in [0, N-1] \tag{5.11}$$

where L represents the size of the L^2-norm estimator, i is the frame index in the time domain, and N is the total number of frames to process.

Additionally, an adaptive threshold method is used to increase the algorithm performance. The method uses a FIFO buffer, containing N_F samples of the energy estimator E_x. Due to the signal periodicity we selected a value for N_F that allows the inclusion of all data needed to compute

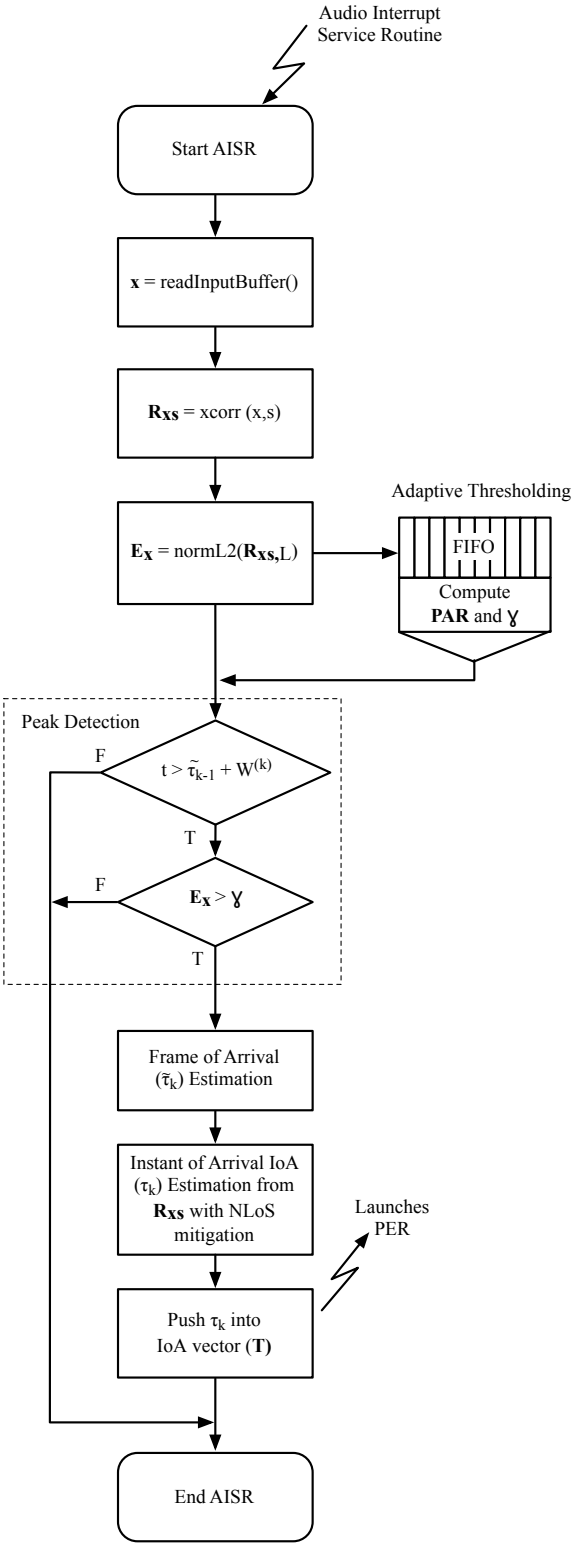


Figure 5.4: Audio Interrupt Service Routine (AISR) algorithm for IoA (τ_k) estimation.

a position estimate using TDMA. With this, we are able to compute the time-varying Peak-to-Average-Ratio (PAR) defined as the ratio between the maximal and the average signal power values, cf. Equation 5.12. The dynamic threshold (γ) is directly obtained from PAR and it is set at 15 dB above PAR, which based on the Receiver Operator Characteristic (ROC) of the matched filter enables a probability of detection of 0.972 for a probability of false alarm of 10^{-3} [86], cf. Figure 3.26.

$$\mathrm{PAR}[i] = \frac{\max_{i \leq i < i+N_F-1} |E_x[i]|^2}{\frac{1}{N_F}\sum_{j=0}^{N_F-1} |E_x[j-i]|^2}, \ \forall i \in [0, N-1] \tag{5.12}$$

Using the decimated energy estimator E_x and the previously computed dynamic threshold γ, a pulse should be valid and detectable based on the combination of the following two main criteria:

1) The decimated energy estimator E_x should be always above γ.

2) A new pulse should only be searched after a $W^{(k)} = S^{(k)} + G^{(k)}$ time interval, counting from the last valid detected pulse.

A valid peak is then searched, to obtain the Frame of Arrival (FoA) for each of the incoming pulses in E_x. Each FoA value is defined by $\tilde{\tau}_k$, where k is the index of the transmitted beacon pulse. Each of these values will be used as a starting point for the search of a more accurate peak detection, i.e. the real Instant of Arrival (IoA) that is defined by τ_k, and it is obtained directly from the R_{xs} signal, as presented on the algorithm in Figure 5.4.

The algorithm runs in the audio interrupt service routine core, and it is called every time a new audio buffer is available for processing. In addition, we have included Non-Line-of-Sight (NLoS) mitigation, to improve the IoA estimation which enhances the pulse detection performance in real situations where multipath occurs, cf. Figure 5.5. Our approach begins with the estimation of the FoA value $\tilde{\tau}_k$ for each incoming pulse only when the two previous criteria 1) and 2) are met, cf. Equation 5.13.

$$\tilde{\tau}_k = \quad \arg\max_i E_x[i], \quad \forall i \in [\tilde{\tau}_{k-1} + W^{(k-1)}, \tilde{\tau}_{k-1} + W^{(k-1)} + L^{(k-1)}] \tag{5.13}$$

When a peak is detected, a more accurate search for the instant of arrival is done directly from R_{xs} in a small interval between $[\tilde{\tau}_k, \tilde{\tau}_k + \delta]$ frames, where δ represents the number of frames (in the example of Figure 5.5 five frames are used) selected to restrict this search in the right side of the neighbourhood of $\tilde{\tau}_k$.

Subsequently, the correspondent IoA search interval in R_{xs} is obtained by $[\tilde{\tau}_k, \tilde{\tau}_k + \delta] \times L$ where L is the size of the L^2-norm estimator. Then the six greater peaks inside the IoA search interval are computed and sorted by the argument value. The algorithm used to detect the first six local

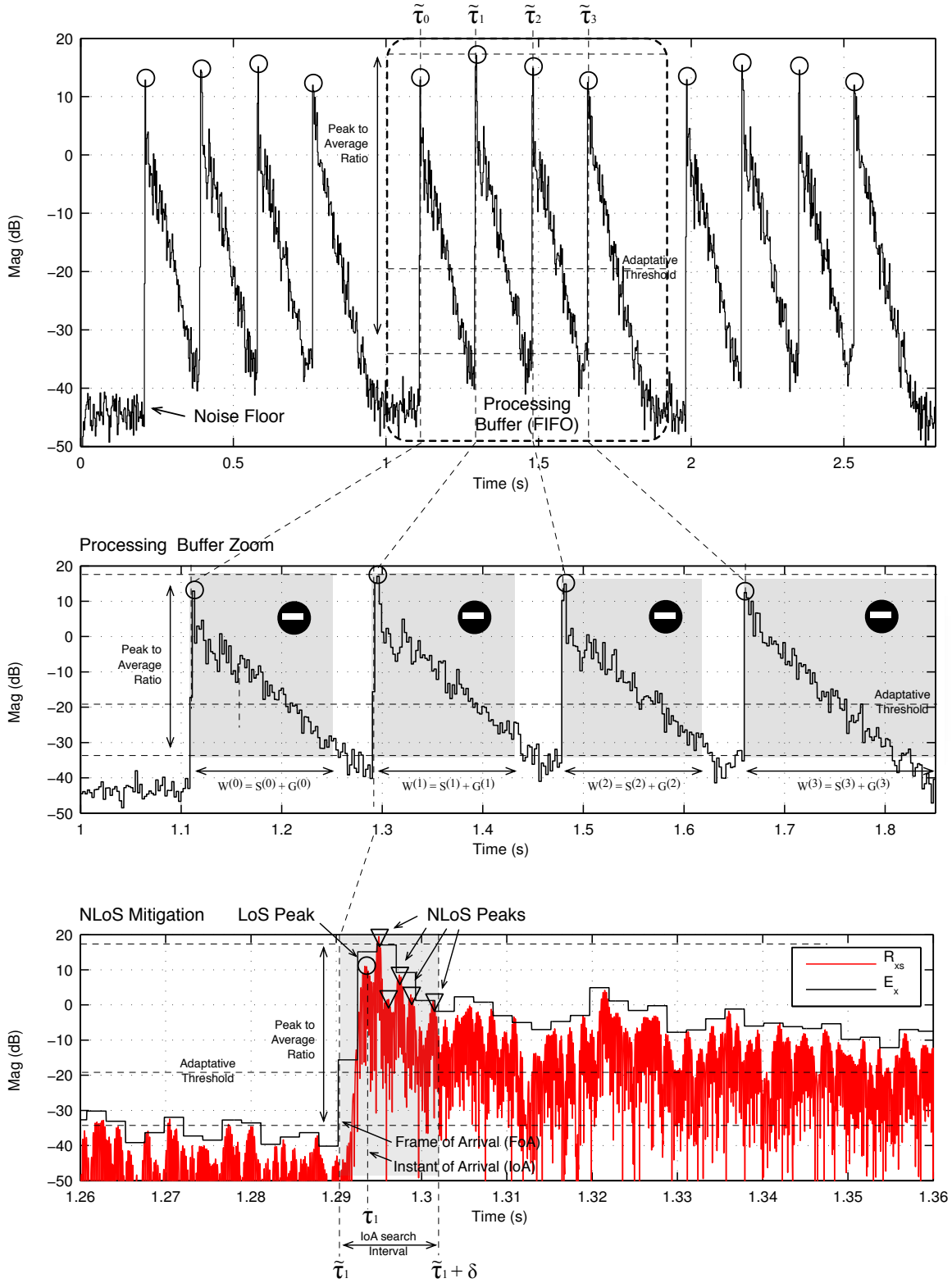


Figure 5.5: Interpretation of the algorithm presented in Figure 5.4, with four acoustic beacons transmitting in TDMA. The signal E_x shows the L2-norm estimation of the signal after the matched filter.

maximums is iterative and starts with the search of the maximum peak in the search interval followed by its removal to prepare the search of the next peak. To remove a peak we considered 0.5 ms for each side of the detected peak. This value was based on the width of the main lobe presented in the autocorrelation function of the proposed signal, cf. Figure 3.28. As we are interested only in the LoS pulse, the peak with lower argument is selected among the other local pulses and assigned to τ_k. The output of this block is a vector $\mathbf{T}$ with a τ_k value per each valid detected pulse.

For the example presented in Figure 5.5, if a simpler criterion based on the argument of the maximum value in the IoA search interval is used, the selected τ_k would point to the first NLoS peak, due to its increased energy when compared to other local peaks. This would introduce the ranging error Δd presented in Equation 5.14, which is far from acceptable in a high accuracy positioning system.

$$\Delta d = c.\Delta t \approx 342.5 \text{ ms}^{-1} \times (1.295 \text{ s} - 1.293 \text{ s}) \approx 69 \text{ cm} \tag{5.14}$$

5.2.3 MT-TDoA Validation

The audio interrupt service routine presented in Figure 5.5 outputs IoA estimates that need to be converted in a valid MT-TDoA set of measurements. This validation process consists in the algorithm presented in Figure 5.6, using a first finite difference operation over the $\mathbf{T}$ data vector to obtain the vector $\mathbf{dT}$ which is then used to evaluate if all of its elements are lower than $T_S^{(k)}$ when $1 \leq k < K - 1$.

Post validation of each $\mathbf{T}$ vector of IoA measurements is needed to obtain a valid difference distance vector $\mathbf{D}$, which is required by the positioning algorithm. To obtain each $D_{k,0}$ value from the $\mathbf{D}$ vector, we need to remove the time slices added by TDMA when computing the MT-TDoA values, and then multiply by the speed of sound. These difference distance $D_{k,0}$ estimates are obtained using Equation 5.15,

$$D_{k,0} = v_{air}[\tau_k - \tau_0 - kW^{(k)}], \ \forall k \in [0, K-1] \tag{5.15}$$

where v_{air} represents the speed of sound, τ_k and τ_0 are a pair of IoA values, $W^{(k)} = S^{(k)} + G^{(k)}$ are directly obtained from TDMA (Figure 4.1) and k is the acoustic beacon index.

Note that for slot synchronization we have used a slightly bigger guard time in the last transmitted pulse $G^{(K-1)}$, which allows the identification of the first valid difference distance $D_{1,0}$ estimate and thus the identification of all subsequent difference distance $D_{k,0}$ estimates for $2 \leq k < K - 1$.

Finally it is necessary to run the position estimation algorithm to obtain the mobile station position. To solve the positioning problem, and since MT-TDoA measurements are always noisy (thermal noise, external acoustic noise, sound velocity changes, etc.), the position estimation can

be seen as an optimization problem. We have opted for finding the position that minimizes the squared error intersection point of the hyperbolas defined by pairs of acoustic beacons.

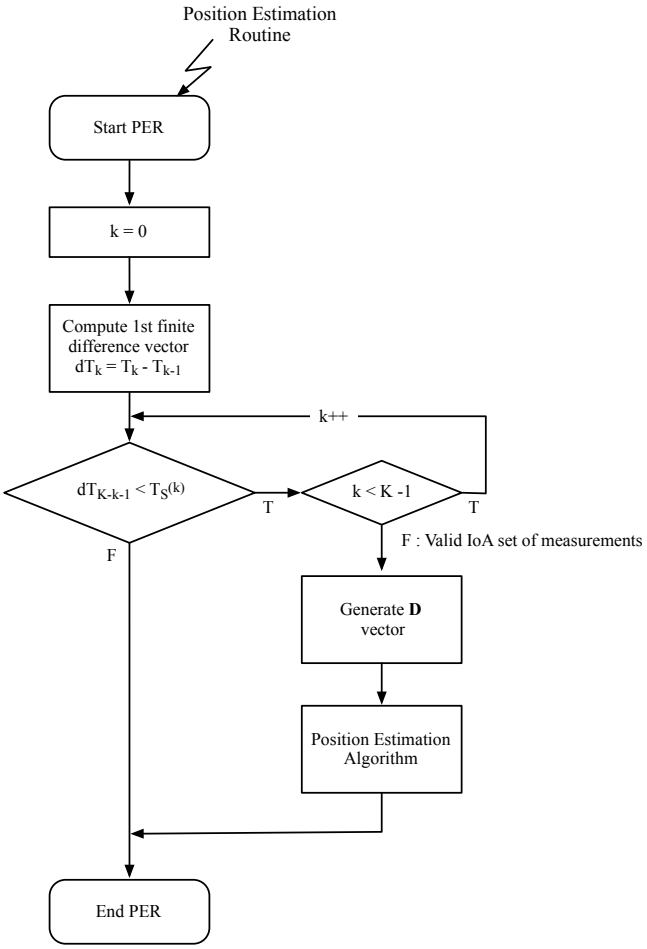


Figure 5.6: MT-TDoA validation routine.

5.3 Hyperbolic Positioning

Hyperbolic positioning systems estimate the position of a target node using the intersection of the hyperbolas that describe the range difference measurements between three or more beacons. These range differences are determined at the target node by measuring the difference in the time of arrival of the incoming pulses multiplied by the speed of sound, as introduced in Equation 5.15. Figure 5.7 illustrates the interpretation of the 2D hyperbolic position estimation for the three hyperbolas with reference to beacon B_0 when four beacons ($K = 4$) are involved.

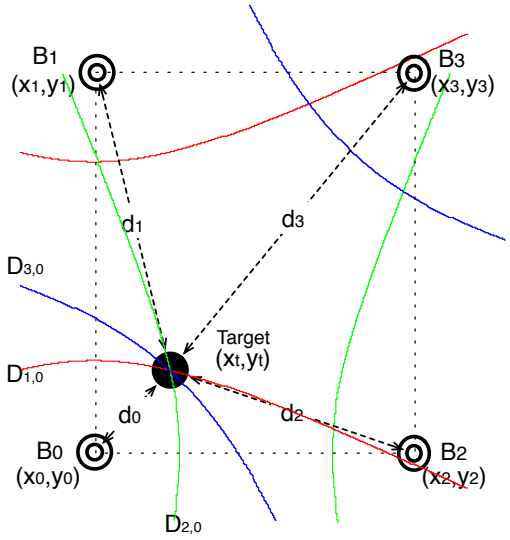


Figure 5.7: 2D hyperbolic positioning interpretation ($K = 4$) with the $D_{i,j}$ hyperbolas represented.

The equations of the hyperbolas presented in Figure 5.7 can be obtained for 2D based on the squared distance equation presented in 5.2, and in the range difference $D_{i,j}$ defined in Equation 5.3, resulting in

$$D_{i,j} = \sqrt{(x_i - x_t)^2 + (y_i - y_t)^2} - \sqrt{(x_j - x_t)^2 + (y_j - y_t)^2}, \quad 0 \leq i < j < K \tag{5.16}$$

In the absence of noise, the unknown position of the target node must lie on the intersection of the three hyperbolas with foci in the beacons positions. The hyperbolas presented in Figure 5.7 are the ones previously described in Equations 5.5, 5.6 and 5.7, and describe the following over-determined system of three non-linear equations,

$$\begin{aligned} D_{1,0} &= \sqrt{(x_1 - x_t)^2 + (y_1 - y_t)^2} - \sqrt{(x_0 - x_t)^2 + (y_0 - y_t)^2} \\ D_{2,0} &= \sqrt{(x_2 - x_t)^2 + (y_2 - y_t)^2} - \sqrt{(x_0 - x_t)^2 + (y_0 - y_t)^2} \\ D_{3,0} &= \sqrt{(x_3 - x_t)^2 + (y_3 - y_t)^2} - \sqrt{(x_0 - x_t)^2 + (y_0 - y_t)^2} \end{aligned} \tag{5.17}$$

Solving the non-linear set of equations previously presented is not easy. Several non-linear and linear approaches can be used to solve the system of equations for (x_t, y_t). The selection of a method to solve the over-determined non-linear system of equations presented in Equation 5.16 depends on several factors, of which we highlight: accuracy, noise statistics information availability, computation efficiency and global convergence. Common approaches used to solve for (x_t, y_t) include the Non-Linear Least Squares (NLLS) and the Linear Least Squares (LLS) estimators. Other

statistical estimators can be used, e.g. Maximum Likelihood (ML), but in general they present a great disadvantage: they require prior knowledge of the noise statistics [9].

The NLLS estimator minimizes the least squares cost function that is directly obtained from Equation 5.16. On the other hand, the LLS methods perform a prior linearization of the equations, and then solve the resulting system of linear equations for (x_t, y_t). Common methods used for linearization include the Squared Differences Method, which will be addressed below, and the Taylor-Series Method, cf. [9] for more details. In the following two subsections we formulate the LLS and NLLS problems in matricial form.

5.3.1 Linear Least Squares Approach

Given that real $D_{i,0}$ measurements are noisy, the position estimation can be seen as an optimization problem. The Squared Differences Method is a simple linearization method where the solution for the set of non-linear equations presented in Equation 5.16 can be obtained with a linear least squares approximation. Consider that beacon B_0 is at the origin $(0, 0)$ (for simplification purposes) and that d_i can be represented by $d_i = d_0 + D_{i,0}$, the Equations presented in 5.17 can be rewritten as,

$$d_0^2 = x_t^2 + y_t^2 \tag{5.18}$$

$$(d_0 + D_{1,0})^2 = (x_1 - x_t)^2 + (y_1 - y_t)^2 \tag{5.19}$$

$$\vdots$$

$$(d_0 + D_{n-1,0})^2 = (x_{n-1} - x_t)^2 + (y_{n-1} - y_t)^2 \tag{5.20}$$

To eliminate the non-linear part we can subtract Equation 5.18 from all other equations, which results in an over-determined set of linear equations. Generally, the number of equations is larger than the number of variables, and the exact solution cannot be obtained (due to the presence of noise), but an estimated solution can be obtained through the least squares technique. The resulting set of linear equations is the following:

$$-x_1 x_t - y_1 y_t = d_0 D_{1,0} + \frac{1}{2}(D_{1,0}^2 - x_1^2 - y_1^2) \tag{5.21}$$

$$-x_2 x_t - y_2 y_t = d_0 D_{2,0} + \frac{1}{2}(D_{2,0}^2 - x_2^2 - y_2^2) \tag{5.22}$$

$$\vdots$$

$$-x_{n-1} x_t - y_{n-1} y_t = d_0 D_{n-1,0} + \frac{1}{2}(D_{n-1,0}^2 - x_{n-1}^2 - y_{n-1}^2) \tag{5.23}$$

Rewriting these equations in matrix form we have

$$\mathbf{H}\mathbf{x} = d_0\mathbf{a} + \mathbf{b} \tag{5.24}$$

where

$$\mathbf{H} = \begin{bmatrix} x_1 & y_1 \\ x_2 & y_2 \\ \vdots & \vdots \\ x_{n-1} & y_{n-1} \end{bmatrix}, \ \mathbf{x} = \begin{bmatrix} x_t \\ y_t \end{bmatrix}, \ \mathbf{a} = \begin{bmatrix} -D_{1,0} \\ -D_{2,0} \\ \vdots \\ -D_{n-1,0} \end{bmatrix}, \ \mathbf{b} = -\frac{1}{2}\begin{bmatrix} D_{1,0}^2 - x_1^2 - y_1^2 \\ D_{2,0}^2 - x_2^2 - y_2^2 \\ \vdots \\ D_{n-1,0}^2 - x_{n-1}^2 - y_{n-1}^2 \end{bmatrix}$$

Finally, the approximate least squares solution ($\hat{\mathbf{x}}$) is given by

$$\hat{\mathbf{x}} = (\mathbf{H}^\mathsf{T}\mathbf{H})^{-1}\mathbf{H}^\mathsf{T}(d_0\mathbf{a} + \mathbf{b}) \tag{5.25}$$

where $\mathbf{H}^\mathsf{T}$ is the transposed $\mathbf{H}$ matrix and $(\mathbf{H}^\mathsf{T}\mathbf{H})^{-1}$ is the pseudo-inverse operator.

In Equation 5.25, the parameter d_0 is unknown. Substituting the intermediate result obtained in Equation 5.18 in Equation 5.25 leads to a quadratic equation in relation to d_0. Solving for d_0 and substituting the positive root back into the least-squares estimation yields the final solution for $\hat{x}$, i.e. the position estimate of the target node. The approach introduced above is the same presented on section 2.3.2.2 but here it was derived for the MT-TDoA case to minimize the squared error intersection point for all the hyperbolas derived from the $D_{i,0}$ measurements.

A common way to solve this optimization problem is to solve it iteratively by computing d_0 using a first guess position followed by substitution in Equation 5.25 to compute the corresponding position estimate. By adjusting d_0 iteratively, a new position estimate is computed, until a given tolerance error is attained. We have implemented this iterative approach to solve the LLS problem, and in Section 5.4 we evaluate its performance.

Since the matrix $\mathbf{b}$ is derived by subtracting Equation 5.18 from all other equations, we can observe that the solution of the LLS problem depends not only on the accuracy of the solution of Equation 5.18, but also on the accuracy of all other equations. If Equation 5.18 contains large errors, the LLS estimator amplifies the measurement noise and the solution can be grossly inaccurate. Several non-linear estimators are known to outperform the LLS estimator, e.g. the NLLS or the ML estimators, for measurements corrupted with Gaussian noise [9].

5.3.2 Non-Linear Least Squares Approach

The NLLS estimator minimizes the least squares cost function directly constructed from Equation 5.16,

$$D_{i,j} \quad = \quad f(x_t, y_t; x_i, y_i, x_j, y_j), \quad 0 \leq i < j < K \tag{5.26}$$

where K is the number of beacons in use, (i, j) is an enumeration that relates two beacons, (x_t, y_t) is the target node position, and (x_i, y_i) represents the ith beacon position. Without loss of generality we can use beacon B_0 as the reference beacon, thus resulting in

$$D_{i,0} \quad = \quad f(x_t, y_t; x_i, y_i, x_0, y_0), \quad 0 \leq i < K \tag{5.27}$$

and in the following cost function

$$\varepsilon(x_t, y_t) \quad = \quad D_{i,0} - f(x_t, y_t; x_i, y_i, x_0, y_0), \quad 0 \leq i < K \tag{5.28}$$

To simplify the notation, let $\mathbf{x} = [x_t, y_t]^T$ and rewriting the cost function $\varepsilon(\mathbf{x})$ in vector notation using a weighted least squares criterion [96, 97] gives,

$$\varepsilon(\mathbf{x}) = (\mathbf{D} - \mathbf{f}(\mathbf{x}))^T \mathbf{R}^{-1} (\mathbf{D} - \mathbf{f}(\mathbf{x})) \tag{5.29}$$

were $\mathbf{D} = [D_{1,0}, \ldots D_{n-1,n-2}]^T$ is the vector with the distance differences obtained from the MT-TDoA measurements and $\mathbf{R} = \text{Cov}(\mathbf{D})$ represents its covariance matrix. The solution defines the minimum variance estimate and the non-linear least squares estimate $\hat{\mathbf{x}}$ is then given by,

$$\hat{\mathbf{x}} = \arg\min_{\mathbf{x}} \varepsilon(\mathbf{x}) \tag{5.30}$$

If a Gaussian noise distribution in the MT-TDoA estimation is assumed, $\hat{\mathbf{x}}$ coincides with the maximum likelihood estimate. With the assumption that

$$\mathbf{D} = \mathbf{f}(\mathbf{x}_0) + \mathbf{e} \tag{5.31}$$

where $\mathbf{x}_0$ is the true position, and noise $\mathbf{e}$ has $\text{Cov}(\mathbf{e}) = \mathbf{R}$, a first order Taylor expansion around the true value gives

$$\mathbf{f}(\mathbf{x}) \approx \mathbf{f}(\mathbf{x}_0) + \mathbf{f}'_{\mathbf{x}}(\mathbf{x}_0)(\mathbf{x} - \mathbf{x}_0) \tag{5.32}$$

From the least squares theory [97] we can obtain

$$\text{Cov}(\hat{\mathbf{x}}) = (\mathbf{f}'_{\mathbf{x}}(\mathbf{x}_0))^{\dagger} \; \mathbf{R}((\mathbf{f}'_{\mathbf{x}}(\mathbf{x}_0))^{\dagger})^T \tag{5.33}$$

were † and ^{T} denote the conjugate and transpose operators, respectively. If Gaussian noise is assumed, Equation 5.33 also defines the Cramer-Rao lower bound, which means that no estimator

can perform better than this bound. Equation 5.33 also lets us extract important information on how a given beacon geometry impacts the accuracy of the final position estimation [98]. In [99], the authors explore the relationship between the Cramer-Rao bound and the Geometric Dilution of Precision (GDOP) matrix, by showing that the GDOP matrix is actually the Cramer-Rao lower bound on estimates of position and bias when the range errors present a Gaussian distribution [99]. For the general non-linear problem presented in Equation 5.30, there exists no closed-form solution and hence, iterative approaches are necessary and shall be described in the following sections.

5.3.2.1 Optimization Methods

In this section we focus on the selection of a NLLS optimization method that will be used to solve Equation 5.30 to obtain a position estimate using an iterative approach. The NLLS iterative optimization algorithms that we will evaluate are presented below in matricial form:

Newton-Raphson : The iterative Newton–Raphson procedure to estimate $\hat{\mathbf{x}}$, is given by

$$\hat{\mathbf{x}}^{k+1} = \hat{\mathbf{x}}^{k} - \mathbf{H}^{-1}(\varepsilon(\hat{\mathbf{x}}^{k}))\ \mathbf{J}^{T}(\mathbf{f}(\hat{\mathbf{x}}^{k}))(\mathbf{D} - \mathbf{f}(\hat{\mathbf{x}}^{k})) \tag{5.34}$$

$$= \hat{\mathbf{x}}^{k} - \mathbf{H}^{-1}(\varepsilon(\hat{\mathbf{x}}^{k}))\ \bigtriangledown(\varepsilon(\hat{\mathbf{x}}^{k})) \tag{5.35}$$

where $\mathbf{H}^{-1}(\varepsilon(\hat{\mathbf{x}}^{k}))$ is the corresponding Hessian matrix operator, $\mathbf{J}(\mathbf{f}(\hat{\mathbf{x}}^{k}))$ is the Jacobian matrix of $\mathbf{f}(\mathbf{x}^{\mathbf{k}})$ computed at $\mathbf{x}^{\mathbf{k}}$ and $\bigtriangledown(\varepsilon(\hat{\mathbf{x}}^{k}))$ is the gradient vector computed at the kth iteration. For more details on how the Hessian matrix and the gradient vector are generated the interested reader is referred to [9].

Gauss-Newton : The iterative Gauss–Newton procedure to estimate $\hat{\mathbf{x}}$, is given by

$$\hat{\mathbf{x}}^{k+1} = \hat{\mathbf{x}}^{k} + \left(\mathbf{J}^{T}(\mathbf{f}(\hat{\mathbf{x}}^{k}))\mathbf{J}(\mathbf{f}(\hat{\mathbf{x}}^{k}))\right)^{-1} \bigtriangledown(\varepsilon(\hat{\mathbf{x}}^{k})) \tag{5.36}$$

$$= \hat{\mathbf{x}}^{k} + \left(\mathbf{A}^{k}\right)^{-1} \bigtriangledown(\varepsilon(\hat{\mathbf{x}}^{k})) \tag{5.37}$$

where $\mathbf{J}(\mathbf{f}(\mathbf{x}^{\mathbf{k}}))$ is the Jacobian matrix of $\mathbf{f}(\mathbf{x}^{\mathbf{k}})$ computed at $\mathbf{x}^{\mathbf{k}}$ and $\bigtriangledown(\varepsilon(\hat{\mathbf{x}}^{k}))$ is the gradient vector computed at the kth iteration. For more details on how the Jacobian matrix and the gradient vector are generated the interested reader is referred to [9].

This algorithm provides a fast convergence and accurate result when a good first guess is used. When a poor initial value is used, or for bad geometric conditions the algorithm results in a rank-deficient, and therefore the matrix A^{k} becomes non-invertible for certain beacon geometries, resulting in an algorithm divergence [100].

Steepest Descent : Contrary to the Gauss-Newton method, the iterative Steepest Descent algorithm is a gradient based procedure with step size μ, and is given by

$$\hat{\mathbf{x}}^{k+1} = \hat{\mathbf{x}}^{k} - \mu^{k} \bigtriangledown(\varepsilon(\hat{\mathbf{x}}^{k})) \tag{5.38}$$

where $\bigtriangledown(\varepsilon(\hat{\mathbf{x}}^{k}))$ is the gradient vector computed at the kth iteration. For more details on how to arrange the Steepest Descent Method, the interested reader is referred to [9].

The easiest way to find a step size μ is to set it to a constant value, $\mu^{k} = \mu$ for all iteration steps. Note that this is a suboptimal solution. The optimum step size for each iteration step k can be determined using an optimum line search procedure which is given by the following non-linear one-dimensional optimization problem [100],

$$\mu^{k} = \arg\min_{\mu} \left(\mathbf{x}^{k} - \mu \bigtriangledown (\varepsilon(\hat{\mathbf{x}}^{k})) \right). \tag{5.39}$$

Due to the high computational effort need for evaluation of the variable step size version of this method we will only consider the evaluation performance of this method using a constant step size. Main drawbacks of this method are the possibility for running into local minima, and slow convergence in the final iteration steps [100].

Levenberg–Marquardt : The Levenberg–Marquardt algorithm is normally used to circumvent the problems of the Gauss–Newton and Steepest Descent algorithms, namely, robustness and slow convergence, respectively. The Levenberg–Marquardt method, based on a damped approach to the Gauss–Newton, is given by

$$\hat{\mathbf{x}}^{k+1} = \hat{\mathbf{x}}^{k} + \left(\mathbf{A}^{k} + \lambda^{k} \mathbf{I}_{2} \right)^{-1} \bigtriangledown(\varepsilon(\hat{\mathbf{x}}^{k})) \tag{5.40}$$

where $\mathbf{A}^{k}$ is the same matrix previously introduced in the Gauss-Newton algorithm, λ is the damping factor, $\mathbf{I}_{2}$ is the 2×2 identity matrix, and $\bigtriangledown(\varepsilon(\hat{\mathbf{x}}^{k}))$ is the gradient vector computed at the kth iteration.

The damping parameter λ^{k} makes sure that matrix $\mathbf{A}^{k}$ can always be inverted, which was not the case on the Gauss-Newton algorithm, thus resulting in a more robust implementation [100]. For $\lambda^{k} = 0$ we obtain the Gauss-Newton solution, and for $|\lambda^{k}| \gg 1$ we approximate the Steepest Descent solution. In [100] the authors present a benchmark for the Gauss–Newton, the Steepest Descent and the Levenberg–Marquardt methods. Levenberg–Marquardt results in a fast convergence and it is very robust when the initial guess is inaccurate. For more details on the description and implementation of the algorithm the interested reader is referred to [100].

5.4 Positioning Algorithm Performance Evaluation

As we are focused on an embedded application, three main requirements should be taken into account in the algorithm selection process: fast convergence to optimal solution, computational efficiency and accuracy in position estimation.

Any of the previous LLS and NLLS methods can be used to solve the position estimation problem. To minimize the average number of iterations for algorithm convergence in the interest area, we have opted for using a first guess based in the centroid estimation for a given beacons geometry. The centroid position can be computed with Equation 5.41,

$$(x_C, y_C) = \left(\frac{\sum_{i=0}^{K-1} X_i}{K}, \frac{\sum_{i=1}^{K-1} Y_i}{K} \right) \tag{5.41}$$

were (x_C, y_C) is the centroid position of the area defined by all beacons B_i.

To evaluate the performance of the iterative optimization algorithms previously introduced in Sections 5.3.1 and 5.3.2 we choose to compute the position estimation using all the algorithms in a grid of 6×7 m (Figure 5.8), which is similar to the grid that will be used later in Section 5.6 for experimental validation. In addition, four beacons were considered at positions $B_0 = (0, 0)$, $B_1 = (0, 8.15)$, $B_2 = (7.25, 0.24)$, $B_3 = (7.25, 8.35)$ and a same initial guess was computed with Equation 5.41 yielding the centroid point (3.63, 4.19).

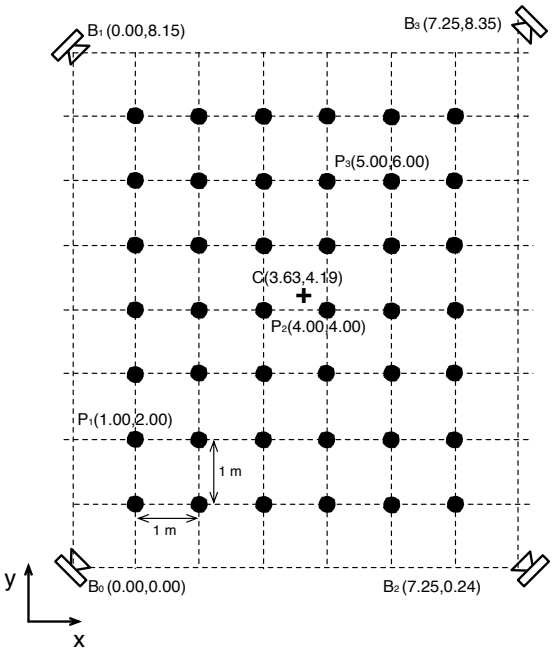


Figure 5.8: Positions used to evaluate the optimization methods previously introduced. The first guess was defined as the geometric centroid of the zone defined by the beacons used in the position estimation process.

Figure 5.9 depicts the result of each of the positioning algorithms in terms of number of iterations needed to converge for a tolerance error of $1e^{-3}$ m, at each distinct position of the grid presented in Figure 5.8. All the algorithms were submitted to the same conditions, i.e. they were fed with a set of difference distances ($D_{k,0}$) computed geometrically for a given target position.

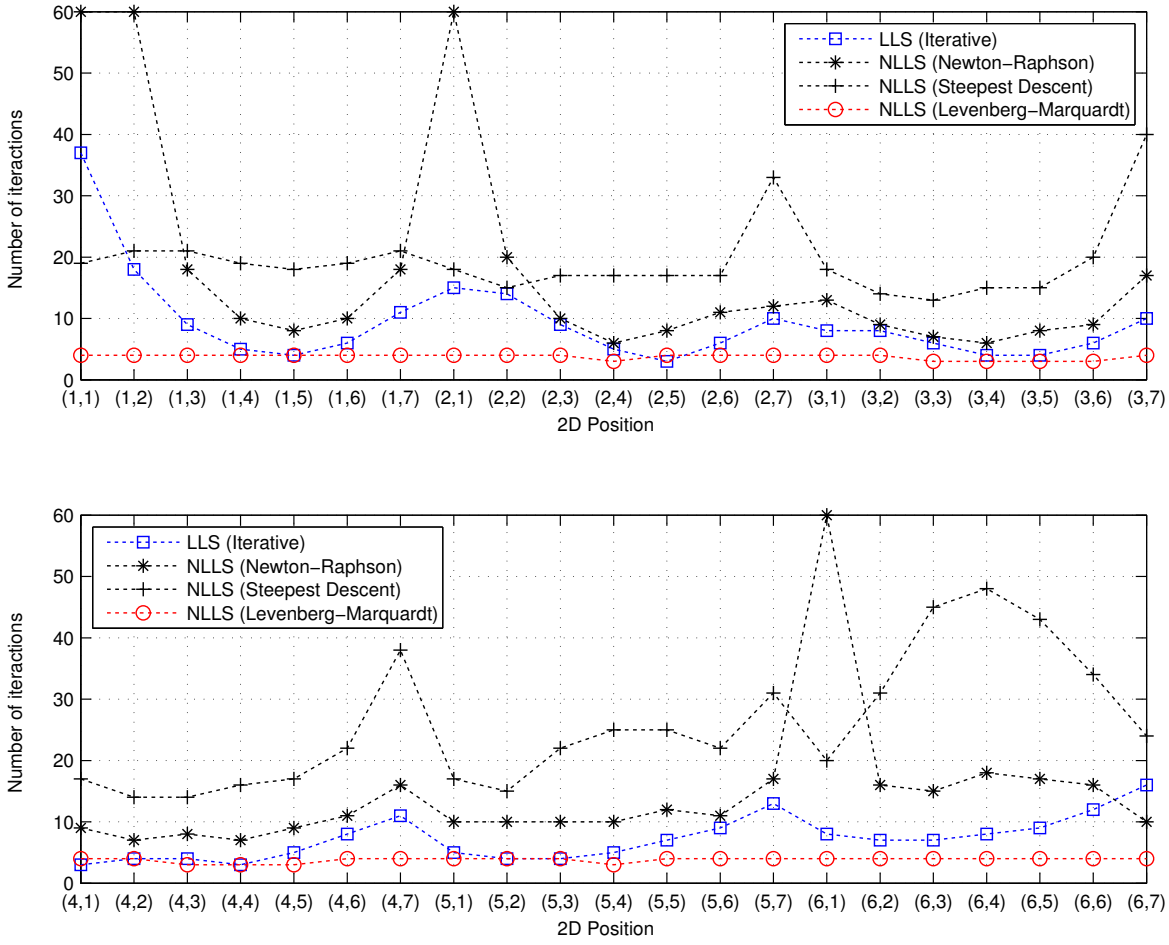
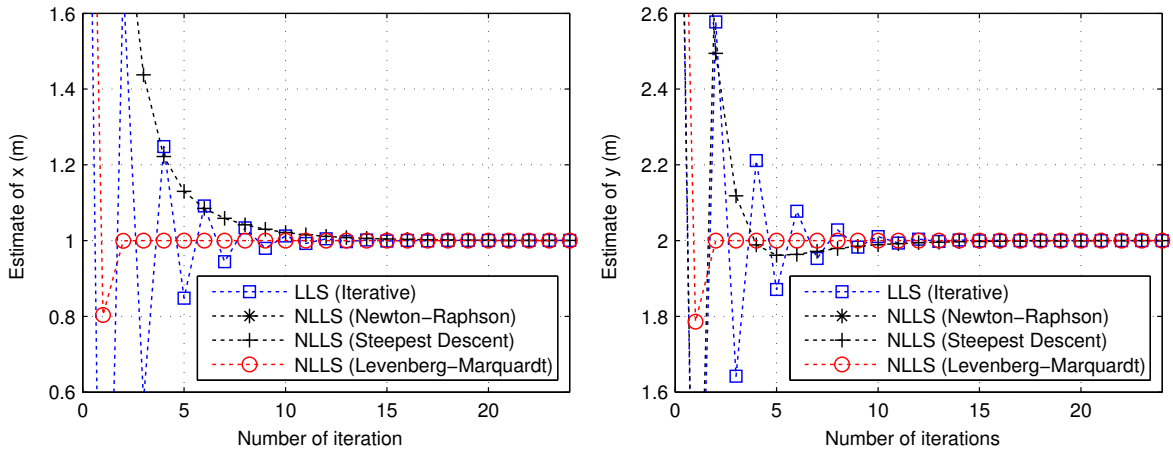


Figure 5.9: Performance of the positioning algorithm in 2D position estimation (x, y) using a tolerance error of 10^{-3} m and the same first guess, i.e. the point given by the geometric centroid of the zone defined by the beacons used in the position estimation process.

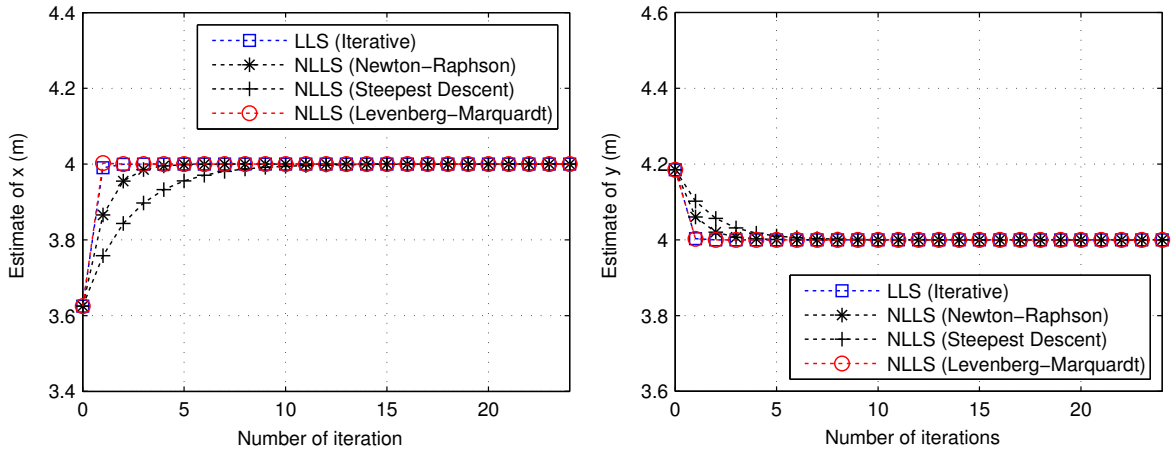
From Figure 5.9 it is possible to observe that the NLLS Newton-Raphson method diverges for positions $(1, 1)$, $(1, 2)$, $(2, 1)$ and $(6, 1)$. On the other hand, the other tested algorithms converge to a tolerance error below $1e^{-3}$ m in less than 50 iterations. The steepest descent algorithm needs on average more iterations to converge due to the fixed step used (in this case $\mu = 0.01$, cf. Equation 5.38).

Figures 5.10, 5.11 and 5.12 present the iterations needed to converge to the optimal solution for the grid positions $(1, 2)$, $(4, 4)$ and $(5, 6)$, respectively.



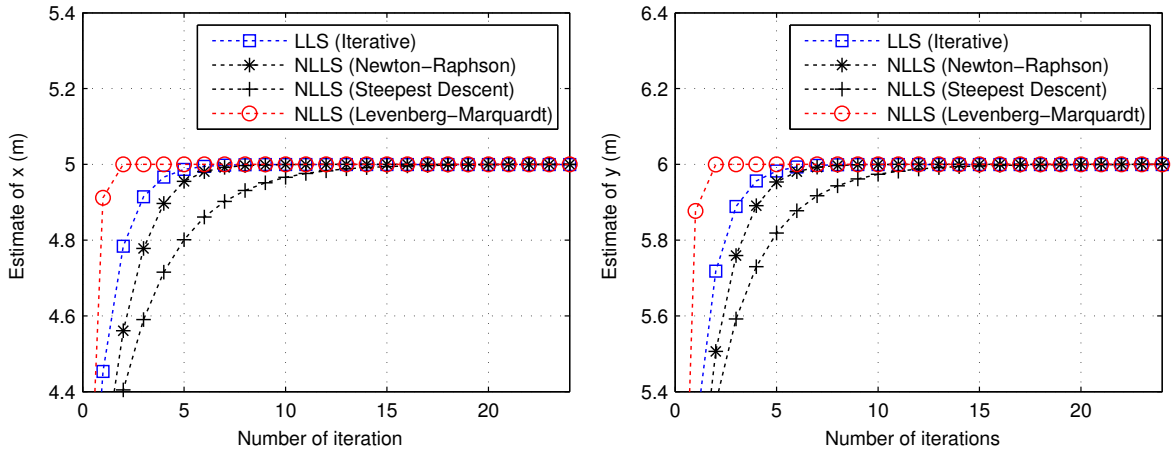
(a) x-coordinate. (b) y-coordinate.

Figure 5.10: Positioning algorithm convergence for the position $(1, 2)$.



(a) x-coordinate. (b) y-coordinate.

Figure 5.11: Positioning algorithm convergence for the position $(4, 4)$.



(a) x-coordinate. (b) y-coordinate.

Figure 5.12: Positioning algorithm convergence for the position $(5, 6)$.

Levenberg–Marquardt is the algorithm that presents the best performance. It converges for a tolerance error below $1e^{-3}$ in less than 5 iterations for all the positions tested in the grid, outperforming all other algorithms examined. It outperforms the well known gradient descent methods, and also avoids dangerous operations with singular matrices [101]. There are two available C/C++ open-source libs that can be used for embedded programming: LEVMAR[1] and LMFIT[2].

Due to real-time and processing constraints imposed by embedded programming, and due to the requirements imposed by the tracking application under development, we have opted for using the LLS iterative approach. This approach shows a good balance between the average number of iterations and complexity. Due to its simplicity we have opted for implementing it in C++, to be integrated in the smartphone app.

5.4.1 The Impact of Geometry / Dilution of Precision

In range-based systems, the beacon-target node geometry heavily influences the position estimation precision [102]. The term Dilution of Precision (DOP) is normally used to estimate the actual contribution of a given geometry to the resulting error in positioning. DOP is normally evaluated using five distinct metrics:

- GDOP: Geometric Dilution of Precision. This metric is normally used to measure the accuracy in position estimation (3D) and in time.
- PDOP: Position Dilution of Precision is normally used to measure the accuracy in position estimation (3D) being also known as spherical DOP.
- HDOP: Horizontal Dilution of Precision. This metric measures the 2D position accuracy in position estimation, namely, Latitude and Longitude.
- VDOP: Vertical Dilution of Precision is normally used to measure the 1D accuracy related to the height of the target node.
- TDOP: Time Dilution of Precision. This metric measures the accuracy in Time at the target node.

The effect of the beacons geometry in the accuracy of a position estimator is called Geometric Dilution of Precision (GDOP) and can be roughly interpreted as the ratio between the position error and the range error [102].

[1] `http://users.ics.forth.gr/~lourakis/levmar/`
[2] `http://apps.jcns.fz-juelich.de/doku/sc/lmfit/`

To compute the metrics previously introduced we start with the derivation of the equations that determine the target node position (x_t, y_t, z_t) and the clock bias distance equivalent d_τ for a given geometric configuration of the beacons [103]:

$$d_i = \sqrt{(x_i - x_t)^2 + (y_i - y_t)^2 + (z_i - z_t)^2} + d_\tau + e_i, \quad i = 0, 1, \ldots K-1 \tag{5.42}$$

where (x_i, y_i, z_i) are the known coordinates of the ith beacon, K is the total number of beacons, d_τ is the clock bias distance equivalent, and e_i are the measurement errors, which are assumed to be the same for each beacon and can be represented by a zero-mean Gaussian noise process with variance σ^2. In this case, the number of unknowns are four ($\mathbf{x} = [x_t \; y_t \; z_t \; \tau]^T$ in matricial form), and at least four non-colinear beacons are required to solve Equation 5.42.

The estimated ranges $\hat{d}_i$ can then be determined based on the current position estimate $(\hat{x}_t, \hat{y}_t, \hat{z}_t)$ and the clock bias distance equivalent $\hat{d}_\tau$ estimate,

$$\hat{d}_i = \sqrt{(x_i - \hat{x}_t)^2 + (y_i - \hat{y}_t)^2 + (z_i - \hat{z}_t)^2} + \hat{d}_\tau \quad i = 0, 1, \ldots K-1 \tag{5.43}$$

From Equation 5.43 we can obtain the partial derivatives with respect to the unknown variables,

$$\frac{\partial \hat{d}_i}{\partial \hat{x}} = \frac{x_i - \hat{x}_t}{\sqrt{(x_i - \hat{x}_t)^2 + (y_i - \hat{y}_t)^2 + (z_i - \hat{z}_t)^2}} \tag{5.44}$$

$$\frac{\partial \hat{d}_i}{\partial \hat{y}} = \frac{y_i - \hat{y}_t}{\sqrt{(x_i - \hat{x}_t)^2 + (y_i - \hat{y}_t)^2 + (z_i - \hat{z}_t)^2}} \tag{5.45}$$

$$\frac{\partial \hat{d}_i}{\partial \hat{z}} = \frac{z_i - \hat{z}_t}{\sqrt{(x_i - \hat{x}_t)^2 + (y_i - \hat{y}_t)^2 + (z_i - \hat{z}_t)^2}} \tag{5.46}$$

$$\frac{\partial \hat{d}_i}{\partial \hat{d}_\tau} = -1 \tag{5.47}$$

resulting in the following unit vectors $(\hat{u}_i)$ that relate the target node and the ith beacon:

$$\hat{u}_i = \left\langle \frac{(x_i - \hat{x}_t)}{\hat{d}_i}, \frac{(y_i - \hat{y}_t)}{\hat{d}_i}, \frac{(z_i - \hat{z}_t)}{\hat{d}_i}, -1 \right\rangle \tag{5.48}$$

where $(\hat{x}_t, \hat{y}_t, \hat{z}_t)$ denotes the estimated position of the target node, (x_i, y_i, z_i) denotes the position of the ith beacon, and $\hat{d}_i$ denotes the estimated distance between the ith beacon and the target node,

cf. Equation 5.43. A visibility[3] matrix H [104] can then be formulated based on the partials derivatives with respect to the unknown variables, i.e. using the unit vectors presented in Equation 5.48 as:

$$H = \begin{bmatrix} \frac{(x_0 - \hat{x_t})}{\hat{d_0}} & \frac{(y_0 - \hat{y_t})}{\hat{d_0}} & \frac{(z_0 - \hat{z_t})}{\hat{d_0}} & -1 \\ \frac{(x_1 - \hat{x_t})}{\hat{d_1}} & \frac{(y_1 - \hat{y_t})}{\hat{d_1}} & \frac{(z_1 - \hat{z_t})}{\hat{d_1}} & -1 \\ \frac{(x_2 - \hat{x_t})}{\hat{d_2}} & \frac{(y_2 - \hat{y_t})}{\hat{d_2}} & \frac{(z_2 - \hat{z_t})}{\hat{d_2}} & -1 \\ \frac{(x_3 - \hat{x_t})}{\hat{d_3}} & \frac{(y_3 - \hat{y_t})}{\hat{d_3}} & \frac{(z_3 - \hat{z_t})}{\hat{d_3}} & -1 \end{bmatrix} \tag{5.49}$$

The first three elements of each row of H are the components of the unit vector $(\hat{u_i})$. Next, a matrix A can be formulated as the inverse matrix of H, and can be used to obtain the DOP metrics previously introduced [103]:

$$A = \left(H^T H\right)^{-1} \tag{5.50}$$

being the elements of A given by

$$A = \begin{bmatrix} \sigma_x^2 & \sigma_{xy} & \sigma_{xz} & \sigma_{xt} \\ \sigma_{xy} & \sigma_y^2 & \sigma_{yz} & \sigma_{yt} \\ \sigma_{xz} & \sigma_{yz} & \sigma_z^2 & \sigma_{zt} \\ \sigma_{xt} & \sigma_{yt} & \sigma_{zt} & \sigma_t^2 \end{bmatrix} \tag{5.51}$$

where σ_x^2, σ_y^2 and σ_z^2 are the variances of the x, y and z coordinates of the target node position estimate, and σ_t^2 is the variance of clock offset estimate [102]. With these variance values we can compute the previously introduced DOP metrics [103]:

$$\text{GDOP} = \sqrt{\sigma_x^2 + \sigma_y^2 + \sigma_z^2 + \sigma_t^2} \tag{5.52}$$

$$\text{PDOP} = \sqrt{\sigma_x^2 + \sigma_y^2 + \sigma_z^2} \tag{5.53}$$

$$\text{HDOP} = \sqrt{\sigma_x^2 + \sigma_y^2} \tag{5.54}$$

$$\text{VDOP} = \sqrt{\sigma_z^2} \tag{5.55}$$

$$\text{TDOP} = \sqrt{\sigma_t^2} \tag{5.56}$$

[3] This name comes from the GPS notation, because it has only information of the satellites that are visible by the GPS receiver when a position estimate is computed.

The resulting DOP values are scalar quantities used to measure the quality of the estimate. Lower DOP values are normally related to a better estimation, i.e. a better a accuracy. Although the DOP metrics previously introduced can be individually computed, they are obtained from the covariance matrix A. This means that the errors are not independent of each other; for example, a high TDOP caused by increased clock error in the beacons will eventually result in increased position errors. Table 5.1 presents the DOP intervals that are commonly used to evaluate the accuracy of the position estimates in GNSS.

In Section 6.2.2 we use the PDOP, HDOP and VDOP metrics to evaluate the performance of the proposed positioning system. In our application, we focus on low density infrastructures, with the number of beacons limited to three or four beacons per room. For a small number of beacons, the optimal beacons position is found to be at the vertices of simple shapes surrounding the mobile station, such as equilateral triangles or squares [105].

DOP	Rating	Description
1	Ideal	Ideal situation. The positioning system always achieves the highest possible precision.
1–2	Excellent	At this confidence level, position estimates are considered to be accurate enough to meet the most sensitive applications requirements.
2–5	Good	This confidence level provides an approximate position estimate. Still reliable for navigation applications.
5–10	Moderate	This confidence level provides less accurate and reliable positioning. Provides a rough estimate of the current location.
10–20	Fair	At this low confidence level, the position estimation is very rough and it is very unreliable.
>20	Poor	At this extremely low confidence level, the performance of positioning system is unacceptable. Unable to locate the target within hundreds of meters.

Table 5.1: Meaning of DOP values used in GNSS, as presented in [106].

5.5 System Prototype

The system prototype is divided in two distinct devices: the acoustic beacon and the mobile device, cf. Figure 5.13. A WSN is used to build an infrastructure of acoustic beacons, with known positions, cf. Figure 5.1. These acoustic beacons can be used as building blocks that can easily be added to an existent infrastructure to meet the scalability criterion. The acoustic beacon was also

designed to include remote configuration via the network. In this way, acoustic pulse parameters (bandwidth, duration and envelope time function) and multiple access features can rapidly be changed to adapt the transmitted acoustic pulses and the multiple access technique to a particular room with specific physical constraints. One acoustic beacon prototype costs approximately 35€, but several simplifications are likely to take place in a production scenario: microphone and preamplifier can be removed, software integration to work with only one μC, and PCB size reduction. With these simplifications we expect that the cost can be reduced to less than 20€.

The mobile device does not need any special hardware to operate. It is only dependent on the smartphone hardware, i.e. the built-in microphone, or any external microphone connected to the headset microphone input. To fulfil the requirements of some pervasive indoor location-aware applications, i.e. augmented/virtual reality, we were forced to add a pair of stereo headphones. This requirement became advantageous because it allowed us to add a microphone on top of the headphones.

Two great advantages were obtained from this decision:

- The microphone became close of the ideal central point of the user head;
- A considerable reduction of the non-line-of-sight situations due to the decrease of the acoustic shadowing effect and the reduction of situations where objects appear between the user head and the room ceiling (where beacons are normally deployed). Due to this, in all experiments we have opted for placing the acoustic beacons at the 2.5 m height plane.

All algorithms (peak detection, IoA estimation, NLoS mitigation and hyperbolic positioning) have been programmed in Objective-C++, using the iOS Accelerate framework for core tasks, such as time domain filtering (convolution) and FFT based operations. The Accelerate framework uses the GPU processor to increase the computational performance taking advantage of its vector and matrix oriented processing capabilities. Each mobile device is able to compute its own position followed by its communication to a remote positioning server through a standard data connection (WiFi or 3G/4G).

Figure 5.13 shows the main views of the developed iPhone app. This app, which we called Akkurate, has been developed to support an audio augmented reality system that is capable of generating virtual binaural sound sources in real-time. The goal is to create the illusion that objects (paintings, sculptures or other works of art in a museum) emit sound. This would allow visitors walking through a particular room, equipped only with headphones and a simple, small and comfortable tracking device (a smartphone), to receive appropriate audio information according to their position.

Figure 5.13: Smartphone-based Mobile Device. a) Mobile Device based in an iPhone 4s running the proposed positioning app (Akkurate). iPhone App user interface views: b) Akkurate app open view. c) Akkurate app main view with actual user position (yellow circle), acoustic beacons (blue circles) and virtual audio sources (red circles) with zone activity delimitation. d) Real-Time debug view mainly for development purposes.

5.6 Experimental Validation

The room used for experimental validation during this work is Lab 4.3.17 of the Department of Electronics, Telecommunications and Informatics of the University of Aveiro. The room has dimensions $9 \times 8 \times 3$ m, concrete walls and ceiling, large windows, linoleum floor and different types of office furniture (desks, chairs, closets, desktop computers and luminaires). Figure 5.14 depicts a detailed floor plan of the room with the beacons added. A grid of 8×9 m with 1 m step was marked in the floor and was used as ground truth in the position evaluation process.

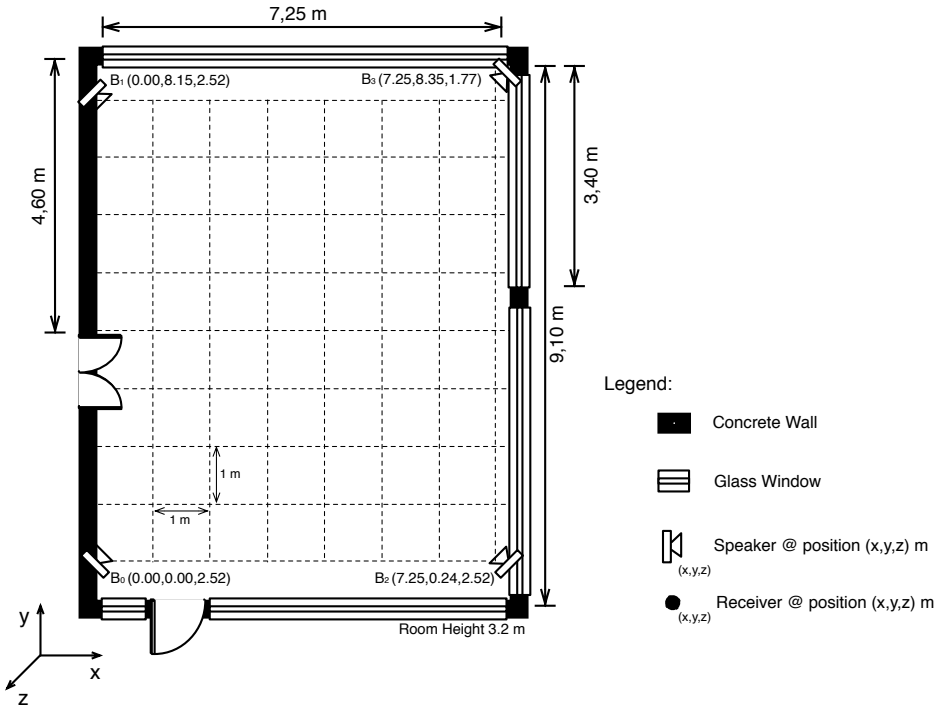


Figure 5.14: Floor plan of the Lab 4.3.17 used for experimental validation.

The room temperature was measured $T = 18°C$, and the sound speed was estimated using the linear model presented in Equation 3.7, thus resulting in $v_{air} = 342.2$ m/s. Additionaly, the room reverberation time at the band of interest was obtained experimentally by observation of the reverberation tail of the room impulse response, i.e. the signal after correlation presented in Figure 5.5, which resulted in a RT_{60} reverberation time [107] of approximately 220 ms.

The non-invasive pulse used for experimental validation is proposed in Section 3.5.1 and its design is based in the time domain weighting approach. This approach was only considered for chirp-based pulses, and for the experiments a pulse with a duration of 10 ms was considered.

The acoustic beacons used in all experimental tests were placed at the following positions:

$$\mathbf{B}_0 : (0.00\ ,\ 0.00\ ,\ 2.52)\ \text{m},$$
$$\mathbf{B}_1 : (0.00\ ,\ 8.15\ ,\ 2.52)\ \text{m},$$

B_2 : (7.25 , 0.24 , 2.52) m,
B_3 : (7.25 , 8.35 , 1.77) m.

Three acoustic beacons (B_0, B_1 and B_2) were placed near the ceiling, all at the same height (coplanar), to reduce NLoS situations in multiple user scenarios. The fourth beacon (B_3) was placed at a lower height to enable the system to obtain 3D position estimates. Two experiments (named A and B) were undertaken to evaluate the overall 2D/3D system performance. Experiment A was designed to obtain a quantitative evaluation of the overall system by comparing the 2D/3D position estimates on a grid of fixed points in the room, and experiment B was undertaken to obtain a qualitative evaluation of the positioning system when a person equipped with a mobile device is in a moving trajectory.

5.6.1 Experiment A: 2D/3D Fixed position estimation

A smartphone running the positioning app was placed at each position marked with a black dot (Figure 5.15), with a constant height of 1.75 m, and one hundred sets of MT-TDoA estimates were obtained for each such position. Figure 5.16a presents the results when the four beacons are used to obtain 3D position estimates, and in Figure 5.16b it is possible to observe the respective 2D projection.

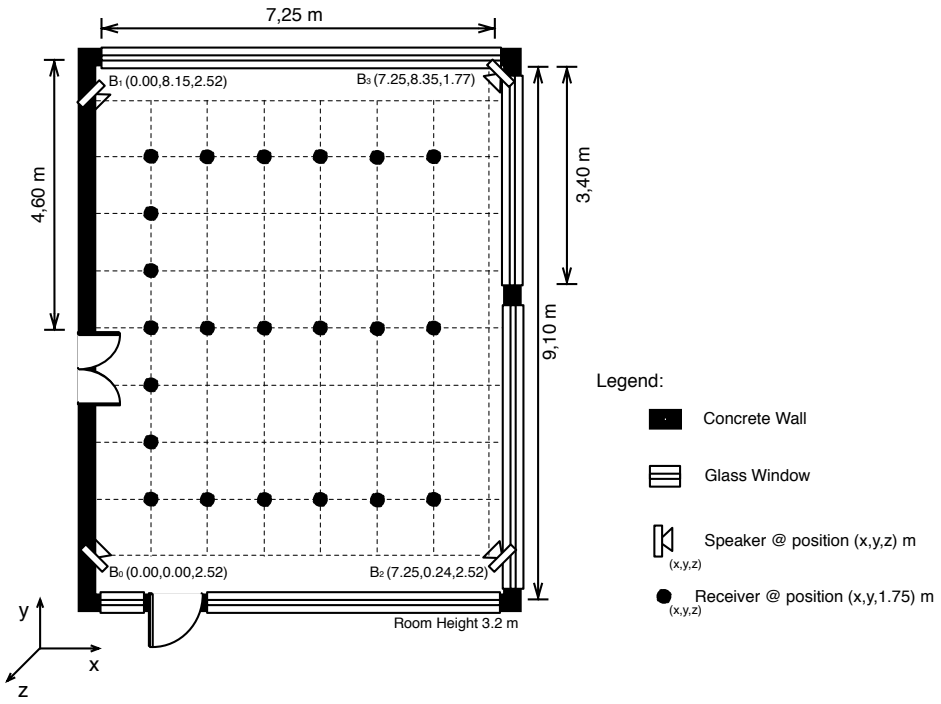
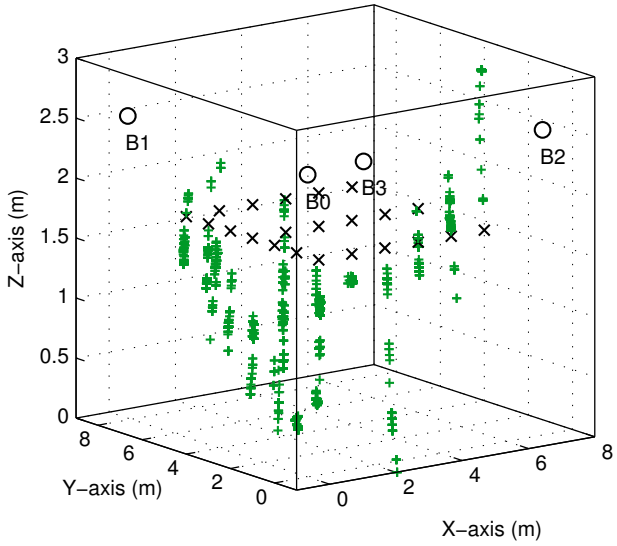
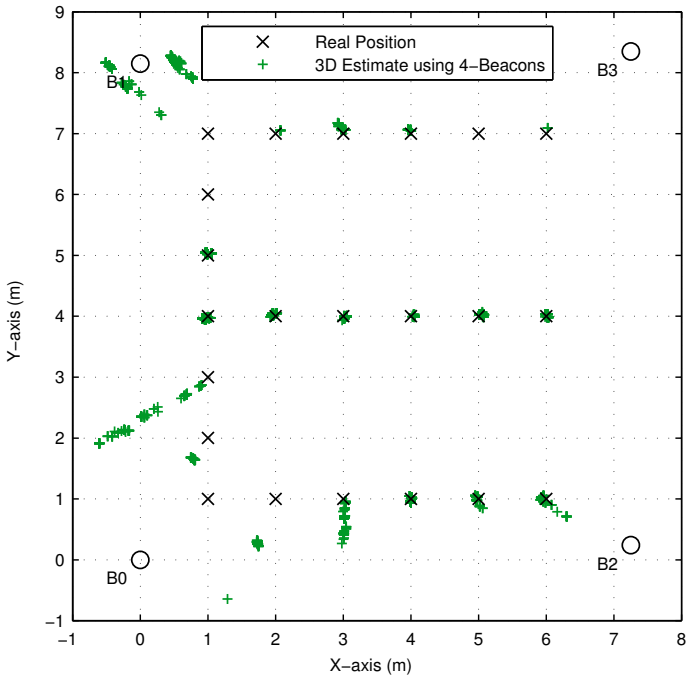


Figure 5.15: Floor plan of the Lab 4.3.17 with experiment A description.

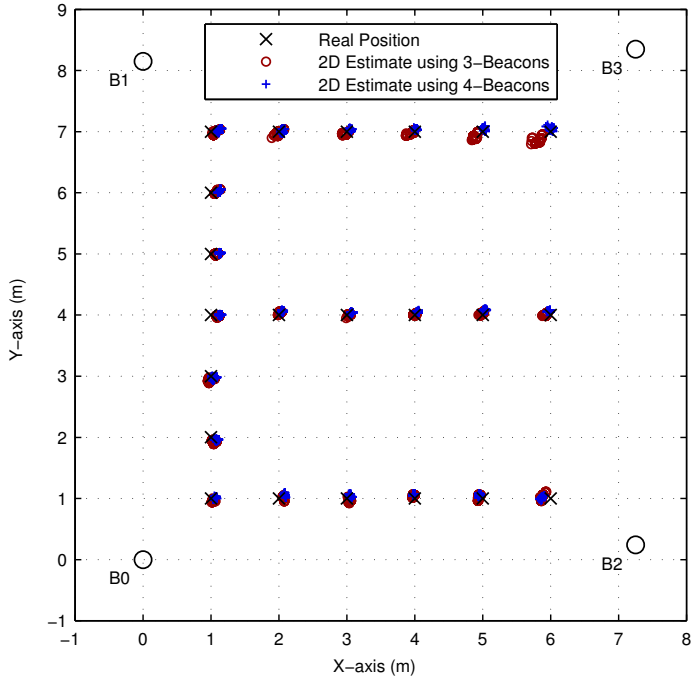


(a) Scatter for the 3D positioning estimates.

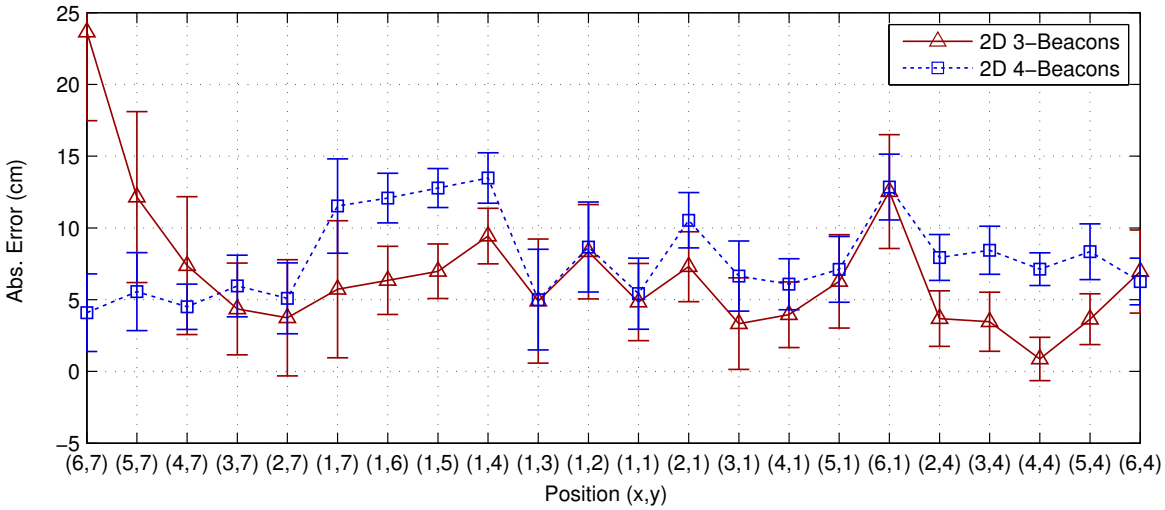


(b) xy projection of the data presented in (a).

Figure 5.16: Experiment A. 3D positioning results with four beacons (B_0, B_1, B_2 and B_3).



(a) Scatter for the 2D positioning estimates.

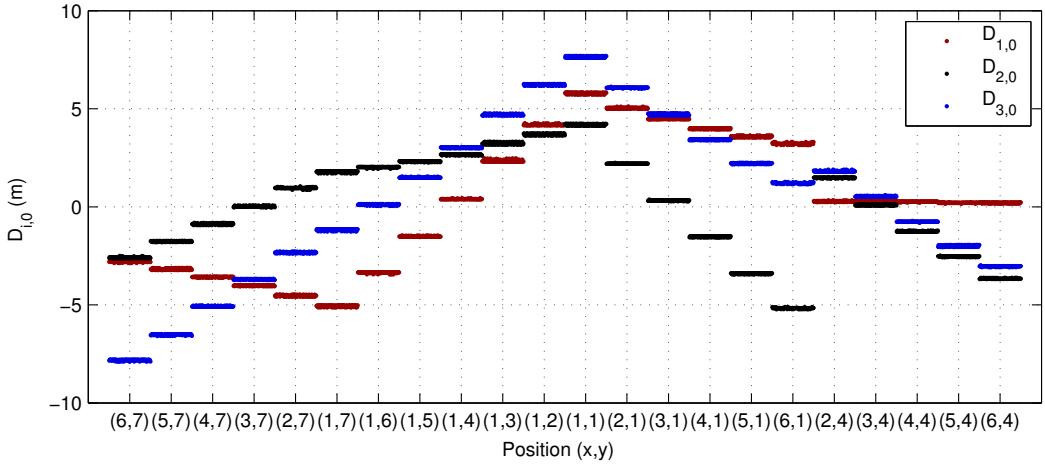


(b) Error statistics for the 2D case.

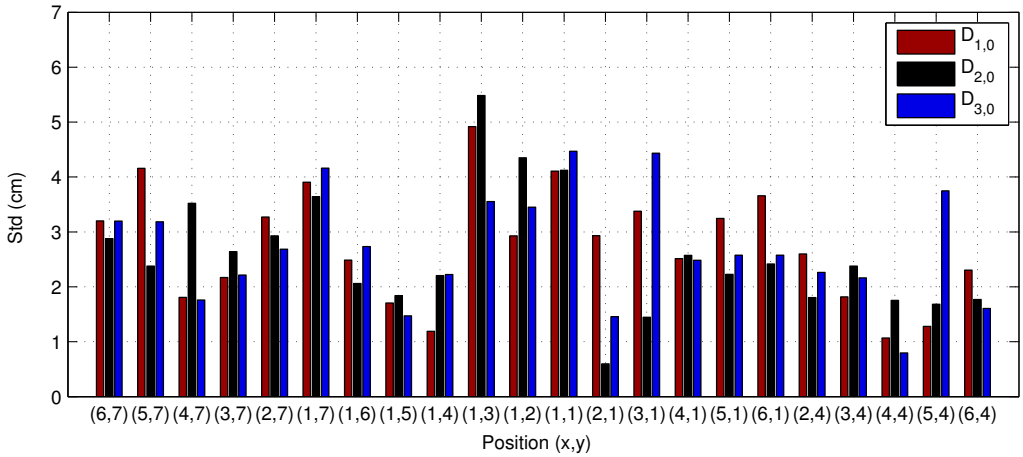
Figure 5.17: Experiment A. 2D positioning results with three and four beacons (B_0, B_1, B_2 and B_3).

Figure 5.17a shows the results when three and four beacons are used to obtain 2D position estimates, and the statistical results (mean absolute error and standard deviation) are presented in Figure 5.17b for all the tested positions. In this case, a systematic error due to non-coplanarity (mobile station at lower height) is introduced, being its impact on error positioning discussed later in Section 6.2, when three or four beacons are used to obtain 2D position estimates. Note that no outlier measurements are present. This can be justified by the fact that all measurements were taken with acoustic noise in the room below 40 dB (SPL).

Figure 5.18a presents all distance differences of arrival $D_{i,0}$ used for position estimation, as described in Figure 5.15. Figure 5.18b shows the respective standard deviation. Note that all measurements present a standard deviation below 6 cm. This error propagates to the final position estimation, and shall be discussed in Section 6.2.



(a) All distance difference of arrival ($D_{i,0}$) resulting measurements.



(b) Standard Deviation.

Figure 5.18: Experiment A. Measurement results. (a) All distance difference of arrival ($D_{i,0}$) measurements. For each position, 120 estimates for each $D_{i,0}$ were obtained and are plotted along the x-axis. (b) Standard deviation for each $D_{i,0}$ in all evaluated positions.

5.6.2 Experiment B: 2D Dynamic position estimation

In this experiment, a qualitative evaluation of the positioning system (2D case) was performed when three and four beacons were used. A moving person with 1.75 m height with the receiver on top of his/her head was used to evaluate the positioning system in two moving trajectories with different speeds, namely one trajectory with average speed of approximately 0.5 m/s and the other with average speed of approximately 1.5 m/s, cf. Figure 5.19.

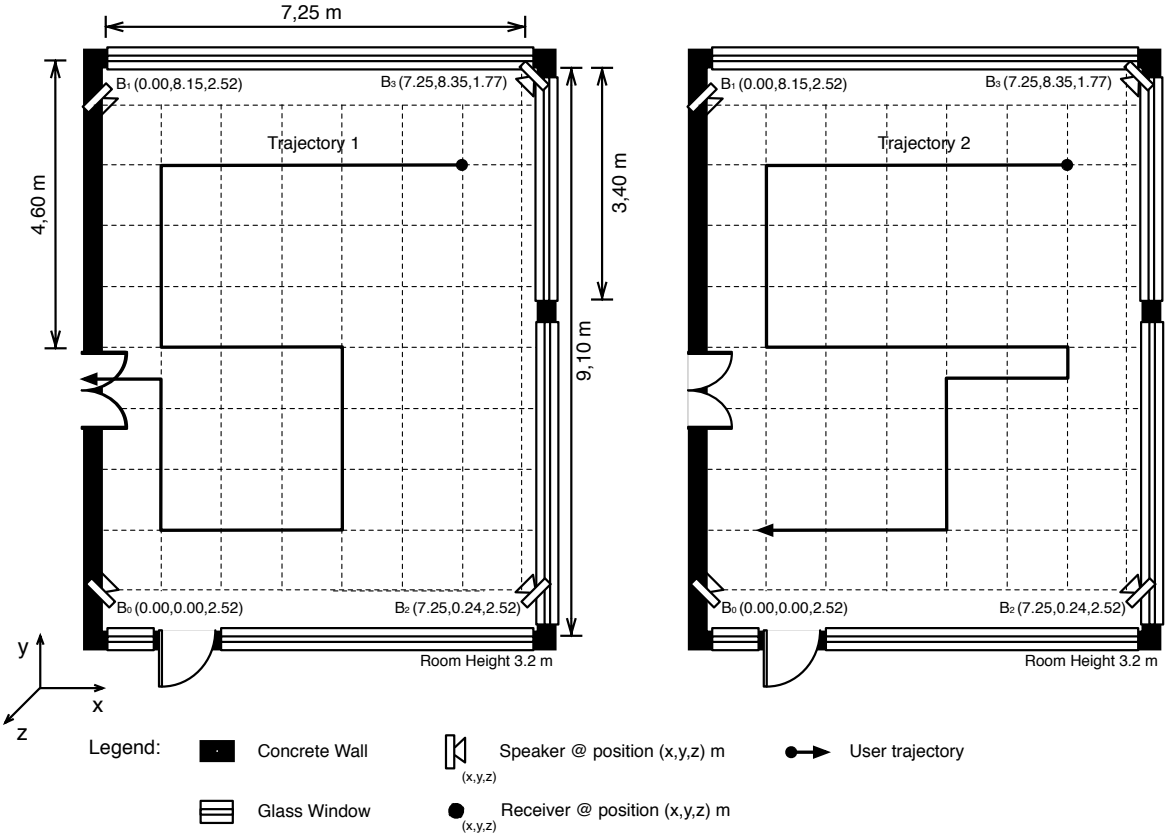
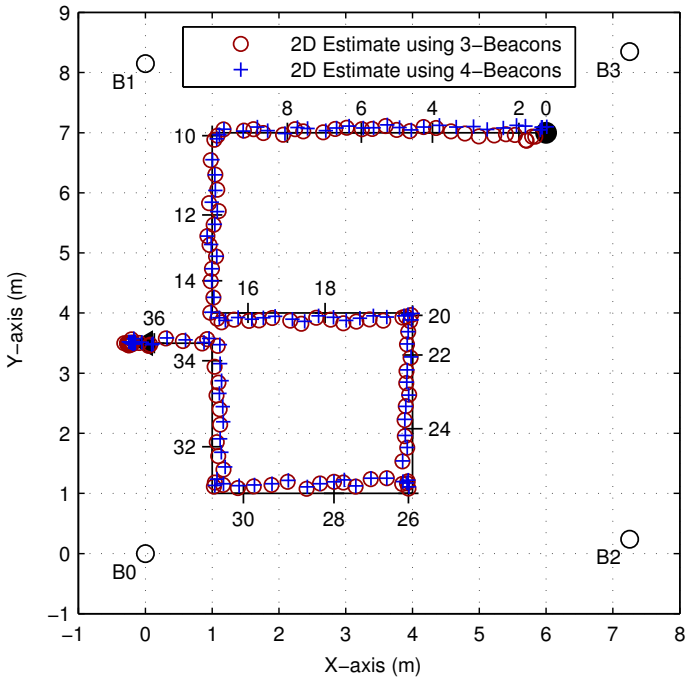
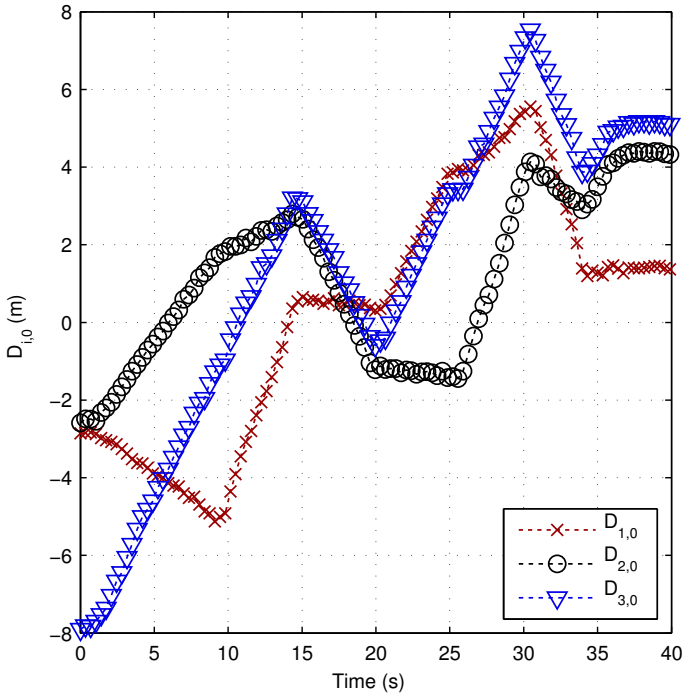


Figure 5.19: Lab 4.3.17 floor plan with experiment B trajectories. In trajectory 1, a user moves with average speed of approximately 0.5 m/s and in trajectory 2 the user moves with average speed of approximately 1.5 m/s.

In this experiment, only a qualitative evaluation can be performed because errors introduced by the natural human body movement cannot be removed from the measured data. Due to this, we cannot perform a correct ground-truth validation based on the trajectories presented in Figure 5.19. For a better interpretation we added the time dimension in seconds along the trajectory line. As a result of the TDMA configuration ($S^{(k)}$=10 ms, $L^{(k)}$=30 ms and $G^{(k)}$=47.5 ms), the lag between the first signal transmission and a position estimation resulted in a mean value of 350 ms with a standard deviation of 23 ms.

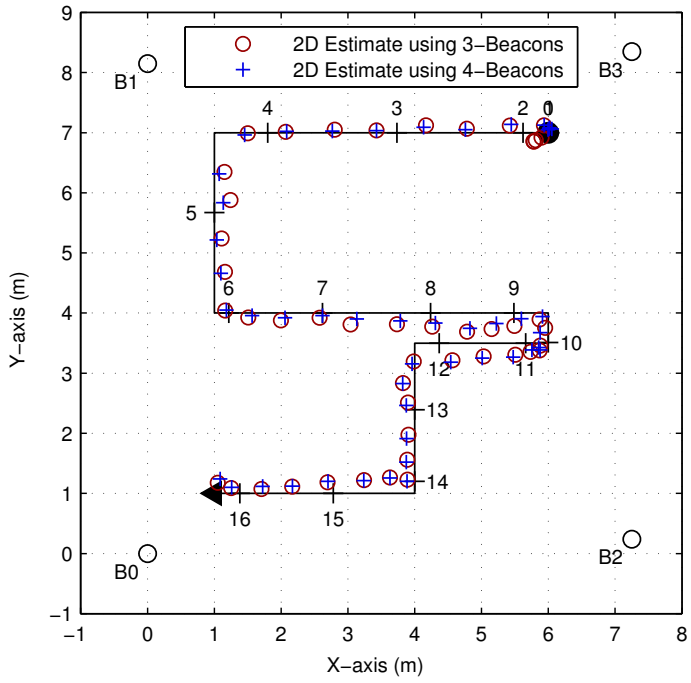


(a)

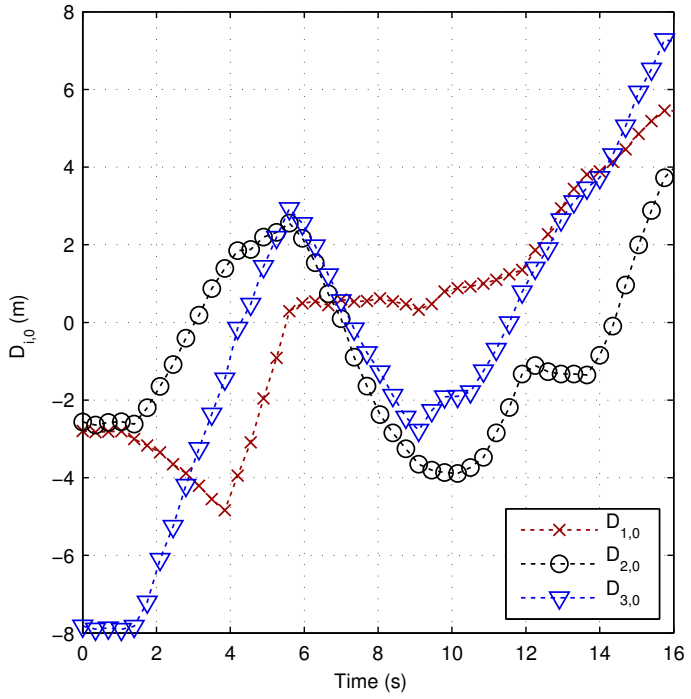


(b)

Figure 5.20: Experiment B. Results for trajectory 1. (a) Trajectory of a user walking at approximately 0.5 m/s. (b) Distance Difference of Arrival ($D_{i,0}$) measurements obtained for trajectory 1.



(a)



(b)

Figure 5.21: Experiment B. Results for trajectory 2. (a) Trajectory of a user walking at approximately 1.5 m/s. (b) Distance Difference of Arrival ($D_{i,0}$) measurements obtained for trajectory 2.

5.7 Chapter Summary

In this chapter we have presented the architecture of a smartphone-based head positioning system named Akkurate. First we introduced the architecture of the proposed system followed by a detailed description of the measurement and positioning approaches. Then, an algorithm to estimate the instant of arrival of the incoming pulses with MT-TDoA validation was presented, and the hyperbolic positioning formulated using the LLS and NLLS methods with a discussion around the selection of the optimization algorithm.

In the end of the chapter the positioning system is evaluated experimentally using an iPhone 4S with two main experiments. The first resulted in a quantitative evaluation of the overall system by the evaluation of the position estimate in a grid of points in a regular room. The second experiment was performed to obtain a qualitative evaluation of the positioning system when a user equipped with a smartphone is in a moving trajectory.

CHAPTER 6

Results and Discussion

In this chapter a discussion on the results and findings is undertaken. The results are presented in two distinct topics. This division comes directly from the structure defined for this thesis and follows the natural sequence used to describe this work. We will start with the discussion of the non-invasive pulse design approaches, followed by a detailed error analysis of the results obtained for the proposed head positioning system.

6.1 Non-Invasive Ranging

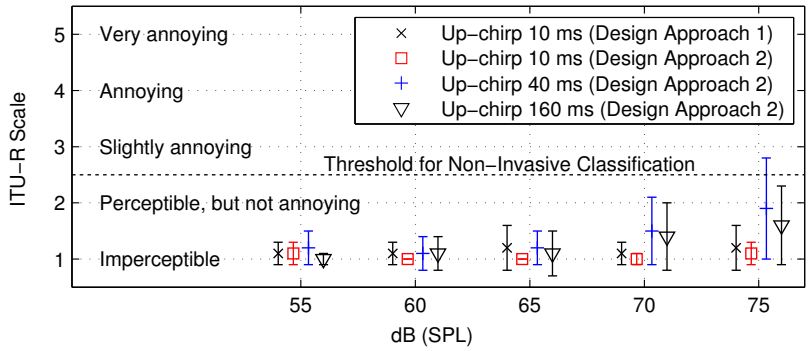
Two approaches were proposed for the design of non-invasive pulses in Chapter 3. The first approach uses time domain weighting to modulate in amplitude the transmitted pulse to smooth its transients and keep the pulse non-invasive. To increase the transmitted power and therefore improve the SNR at the output of the matched filter, an asymmetric weighting window was presented to modulate the transmitted pulse in amplitude. This approach was only tested with linear frequency modulated pulses, i.e. chirps, due to their known and constant frequency rate over time. With this simple technique, we were able to increase the transmitted power by approximately 5 dB and reduce the compression ratio (CR) at the output of the matched filter, cf. Table 3.3.

The second approach uses a perceptual filter designed to reduce the perceptual impact of the pulses described in Section 3.5.2.1. The proposed filter was based on the study presented by Suzuki and Takeshima in 2004 [10], which includes a threshold of hearing contour based on twelve contemporary loudness studies. On this approach we use a yardstick pulse as reference to design several other pulses with the same energy but distinct durations and modulation types.

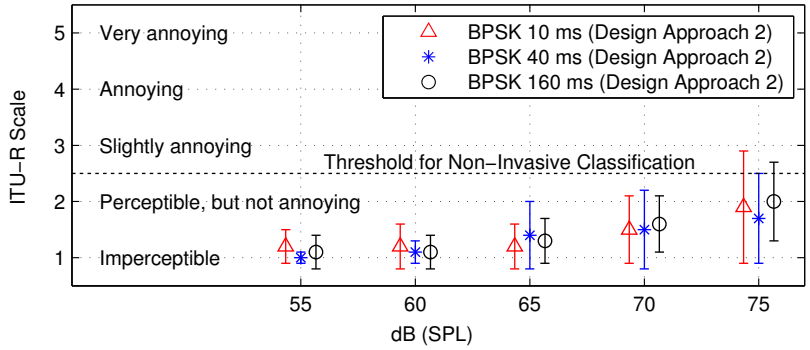
To evaluate both design approaches a perceptual evaluation experiment was performed, cf. Section 3.6. The perceptual evaluation results are compiled in Appendix C and will be analysed and discussed next.

6.1.1 Perceptual Evaluation Analysis

In Figure 6.1 are presented the perceptual evaluation statistical results over all subjects, namely, the mean and standard deviation for the evaluation of the same pulse. The proposed definition of non-invasive audio signals presented in Chapter 1 was generally accepted by the subjects; even when they were able to detect the presence of the stimulus, all agreed that the general classification of the stimulus as non-invasive audio was fairly acceptable.



(a) Chirp-based Pulses.



(b) BPSK-based Pulses.

Figure 6.1: Perceptual evaluation statistical results when all the subjects are considered. Mean and standard deviation when the subjects were evaluating the same pulse.

Due to the hearing threshold variation with the subjects age (Figure 3.14) we have opted for analysing the experimental data in two age intervals: 16-29, and above 30.

Figure 6.2 presents the statistical results relative to the 16-29 years old subjects. From the data it is possible to observe that the perceptual impact is greater for the BPSK-based pulses, which start being perceptible in average for sound pressure levels above 65 dB. On the other hand, the chirp-based pulses start becoming perceptible for sound pressure levels above 70 dB with the exception of the two 10 ms duration chirp-based pulses that remain almost imperceptible for all SPL levels.

Figure 6.3 shows the statistical results relative to the subjects of age over 30 years. In this case, the subjects marginally perceive the pulses for higher sound pressure levels. For the chirp-based pulses, the pulse with the duration of 160 ms is the one that presents a residual impact for sound pressure levels above 65 dB. Moreover, if we consider the BPSK-based pulses only at 75 dB of sound pressure, the subjects were able to detect the presence of the pulse.

The data presented in Figures 6.2 and 6.3 can be related with the theory of loudness perception presented in Section 3.2.2. Although the loudness perception dependence on the stimulus duration has been discussed for frequencies in the range 1-4 kHz, the resulting temporal integration tends to be more noticeable for higher frequencies [65]. This means that the loudness sensation for the frequency band we are using, i.e. between 16-24 kHz should be greater than the values presented in Figure 3.17. Moreover, if we consider the theory behind the loudness sensation of complex signals presented in Section 3.2.2, it only considers stimulus with multiple tones around a tone of 1 kHz [77], cf. Figure 3.18. We have not found any evidence in the literature, about the loudness spectral summation for frequencies in our band of interest, i.e. 15-25 kHz.

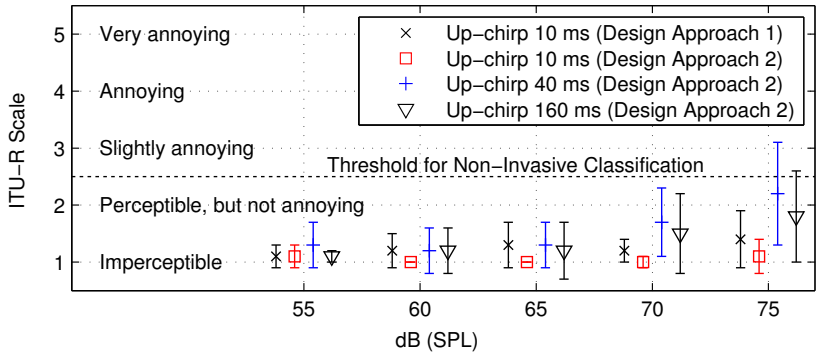
Additionally, if we consider that the hearing threshold in the upper frequency (Figure 3.14), decreases with age, phenomenon known as "presbycusis", then the experimental data correlates with the literature as expected, with all designed pulses almost inaudible for 30 years old and above subjects, for sound pressure levels below 75 dB.

6.2 Position Estimation Error Contributions

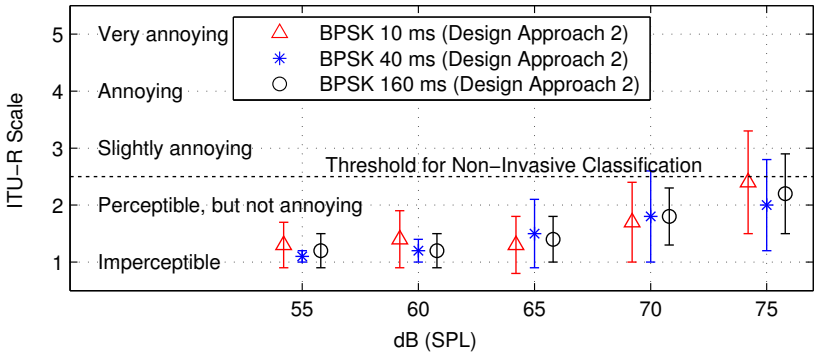
In this section we analyse in detail the two major contributions to the final position estimation error: (1) MT-TDoA Estimation Error and (2) Geometry Error, mainly related to the network density and geometry, a problem known in the literature as DOP, cf. Section 5.4.1.

6.2.1 MT-TDoA Estimation Error

The MT-TDoA estimation has several error contributions. Main contributions are related to beacon synchronization and multipath propagation. The former is directly related with the offset sync error observed back in Section 4.3.3 and the latter is a well known phenomenon that heavily affects the LoS peak estimation. Additionally, the speed of sound uncertainty also affects the final distance difference estimation. These three main sources of error contribute independently to the distance difference estimation and are discussed next.

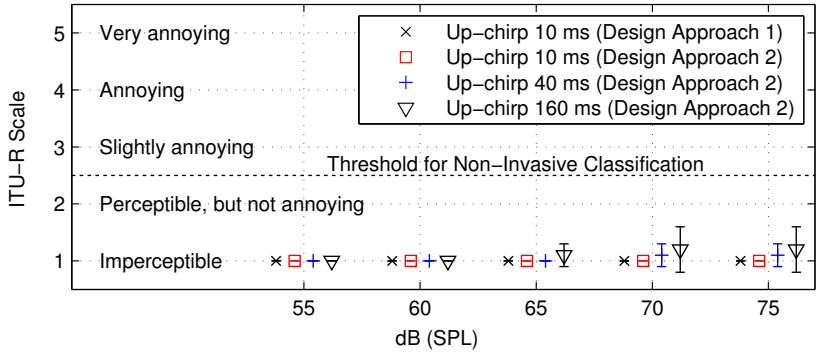


(a) Chirp-based Pulses.

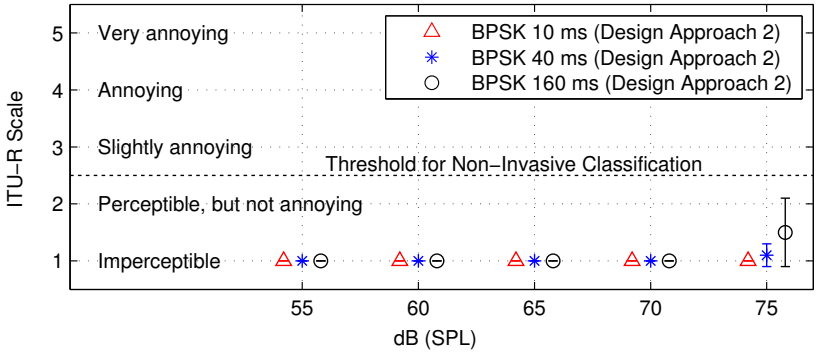


(b) BPSK-based Pulses.

Figure 6.2: Perceptual evaluation statistical results for subjects aged between 16 and 29 years old. Mean and standard deviation when the subjects were evaluating the same pulse.



(a) Chirp-based Pulses.



(b) BPSK-based Pulses.

Figure 6.3: Perceptual evaluation statistical results for subjects over 30 years old. Mean and standard deviation when the subjects were evaluating the same pulse.

In all experiments the temperature was measured using a calibrated digital thermometer and was considered constant during the experimental phase, i.e. $T = 18° \pm 1°$ C. According to Equation 3.7, for the measured temperature the speed of sound is 342.2 $\pm$ 0.6 m/s.

6.2.1.1 Synchronization Error

The synchronization error is mainly introduced by the proposed ATSS protocol, affecting the MT-TDoA estimation process and consequently the distance differences estimation. The proposed protocol was evaluated experimentally using the procedure defined in Section 4.3.3. Figure 4.8g shows the absolute correction value that should be performed over time. From the observation of the linear regression data in Figure 4.8g, we can state that the clock offset sync correction should be performed at a rate of approximately 4 μs/s. Another observation from the data presented in Figure 4.9g is that the absolute offset error is always below 8 μs, for all distinct synchronization periods. The worst performance of the ATSS protocol is for a synchronization period of 1 s, which is justified by the high number of exchanged messages between beacons that result in a processing overhead in the microcontroller unit. Considering the worst case scenario of 8 μs in the synchronization error, and neglecting the error associated with the sound speed estimation, an absolute distance error of approximately 2.7 mm is obtained, which has a reduced impact on the final position estimation. Another important observation that can be inferred from the data presented on Figure 4.9 is that the syntonization process distributes the offset correction over time. This enables the use of long synchronization periods, e.g. 30 s or 60 s, keeping the computational overhead imposed by the use of the ATSS protocol at a low level, thus resulting in a residual power consumption overhead.

6.2.1.2 Peak Detection Error

In Section 3.4 the peak detection error is given in terms of probability of detection with respect to SNR. Figure 3.26 shows that an improve in the detection performance can be obtained increasing the SNR at the output of the matched filter. The peak detector presented in Section 5.2.2 uses a dynamic threshold (γ) that is directly obtained from the PAR and it is set to be 15 dB above PAR, which based on the ROC of the matched filter (Figure 3.26), enables a probability of detection of 0.972 for a probability of false alarm of 10^{-3} [86].

In the presence of noise, a detected pulse always shows jitter in the time domain. The use of a particular sampling rate in the mobile device affects the peak detection due to resolution constraints. Additionally, the PSL and CR figures of merit also impact this jitter variability, when noise is present. The influence of the sampling rate in the peak detection can be quantized, i.e. the distance error for one sample deviation in the peak detection. In experimental validation we have

used a sampling rate of 44.1 kHz in the smartphone. This results in a peak detection ambiguity of 7.8 mm per sample, i.e. 1/44.1 kHz × 342.2 m/s = ≈ 7.8 mm, which also has a reduced impact on the final position estimation.

6.2.2 Geometry Error

In our application case, we focused on low density infrastructures, being the number of beacons limited to three or four beacons per room. Therefore, an optimal placement of these beacons is crucial to obtain the mobile station position with increased accuracy [108]. For a small number of beacons, the optimal beacons position is found to be at the vertices of simple shapes surrounding the mobile station, such as equilateral triangles or squares [105]. During experimental validation, this criterion was only satisfied in the 2D case. Note that the position of beacon B_3 is not optimal due to physical room constraints. The lower the height of B_3, the better should be the position estimator performance.

An optimal placement for B_3 would present a negative height around -5 m, which would be impossible to satisfy. Due to room constraints (room geometry and existing furniture) we were forced to put B_3 at 1.77 m height, which was the lowest value we were able to use. The 3D positioning results presented in Figure 5.16a shows the effects of the geometry problems previously discussed. The non-optimal position of beacon B_3 resulted in a high sensitivity to noise in most of the 3D position estimates which can be confirmed by the spread of the results mainly in the z-coordinate. In addition, for some positions the estimator cannot solve for a valid solution (i.e. inside the room volume) which can be demonstrated by the 2D projection of the 3D estimates for positions (3,7) and (4,4), cf. Figure 5.16b.

Figure 6.4 presents the DOP analysis for the setup used for experimental validation, cf. Section 5.6. As described in Section 5.4.1, the DOP metric gives us important information about the accuracy of the positioning for a given beacon-target node geometry. We have opted for computing all the DOP metrics, i.e. GDOP, PDOP, HDOP, VDOP, TDOP for the positions tested in Experiment A, cf. Figure 6.4. From the data presented in Figure 6.4, and according to the information presented in Table 5.1 it is possible to observe that the only metrics that present a good accuracy are the HDOP and TDOP, i.e. with an average rating between ideal and good. This is in line with 2D results.

On the other hand, the VDOP and consequently the GDOP metrics are dispersed with rating from moderate to poor, which is aligned with the 3D results presented in Section 5.6. In the 3D case the performance is also affected by the position estimation algorithm that diverges for the positions $(4, 4)$ and $(3, 7)$. As we are mainly interested in 2D position estimation, and knowing that for the geometry used in experimental validation we have a bad performance of the VDOP metric (which results in low accuracy in z-coordinate), we have opted for evaluating the proposed system only for 2D.

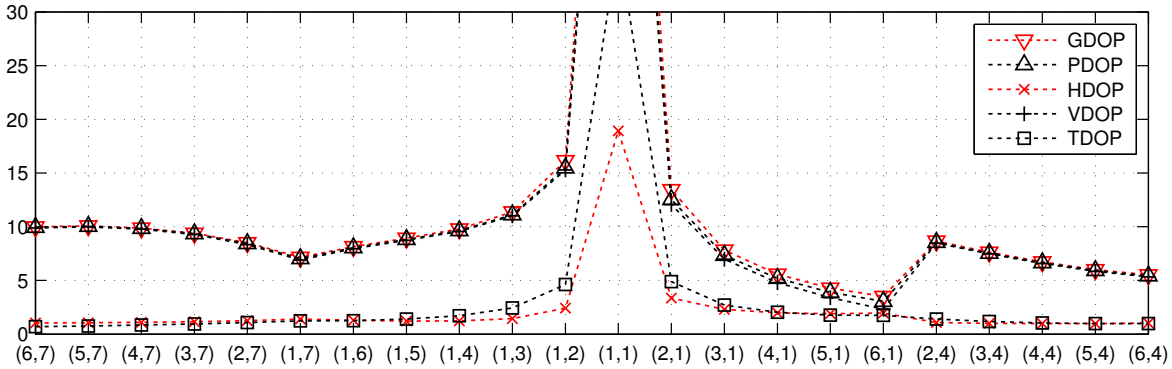


Figure 6.4: DOP analysis of the experimental setup used in experiment A.

6.2.2.1 Non-coplanarity Contribution to 2D Error

In 2D, the network geometry affects the absolute error mainly due to the non-coplanarity of the mobile station with the beacons plane, due to the different height values that it can present, cf. Figure 6.5. In the coplanar case no error is introduced by the geometry and a valid solution can be obtained by constraining the search of the solution by the positioning algorithm to fit inside the room area. The 2D projected distances can be computed by $d'_k = \sqrt{d_k^2 - h_c^2}$ where h_c can be obtained directly from $h_c = h_b - h_m$. In the situation presented in Figure 6.5 the measured distances d_k are always greater than their 2D projection d'_k. In this case, these larger measured distances affect the final position estimation, contributing differently to the overall absolute error, depending on the position and height of the mobile station.

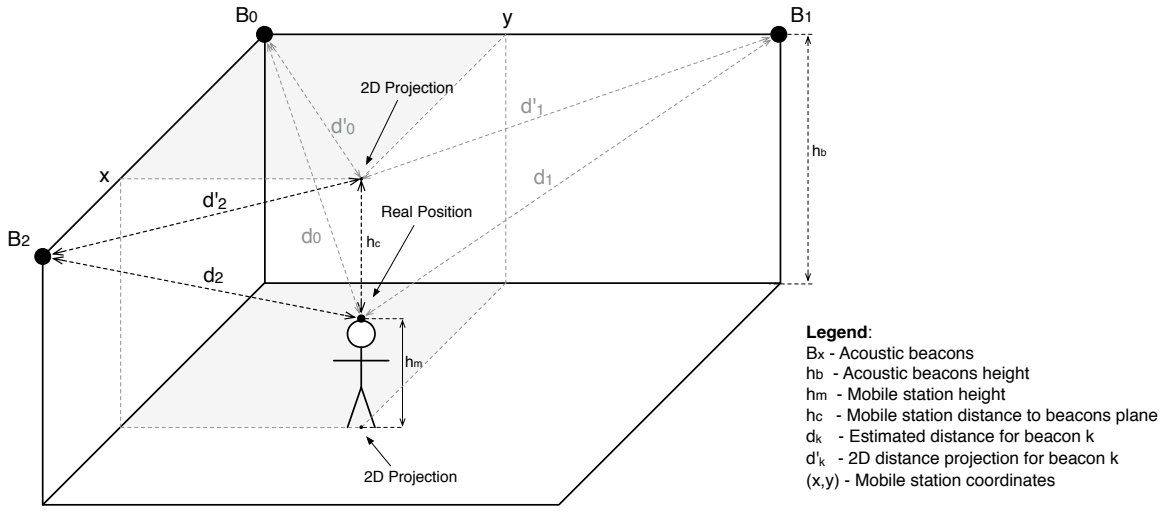
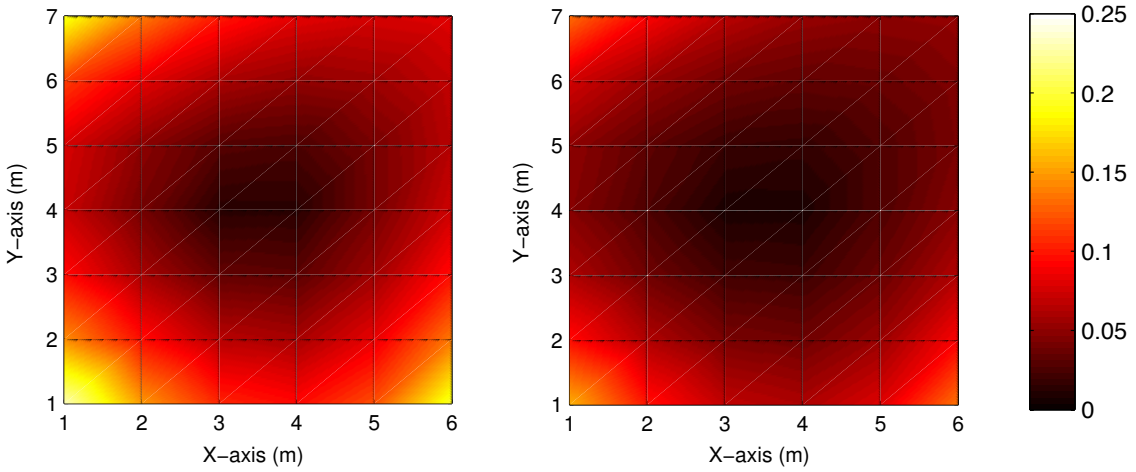
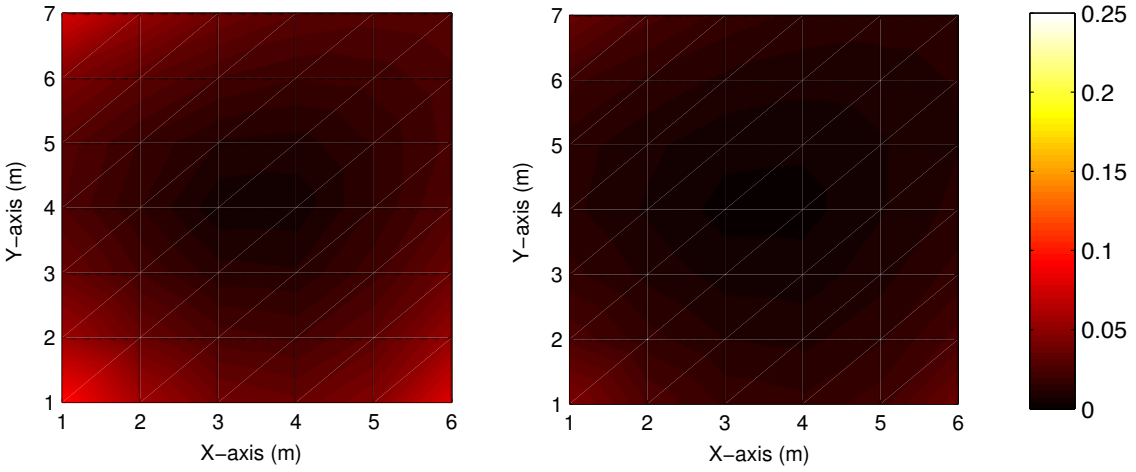


Figure 6.5: Illustration of the 2D projection obtained from a target node with variable height.

In Figure 6.6 is presented the error map obtained by simulation when the positioning algorithm is fed with noise-free TDoA measurements for the situation presented in experiment 1 (three beacons) for different mobile station height values. In these simulations the receiver height is changed between 1.25 m and 2 m to obtain a 2D grid of the absolute error. The absolute 2D error increases for peripheral positions near the acoustic beacons and considerably reduces in the central zone. This can be geometrically justified by the lack of height information, which leads to distance measurements with values always above the real value and with greater impact near the acoustic beacons when the mobile station height deviates from the acoustic beacons plane. This analysis is correlated with the 2D results (Figure 5.17b).



(a) $h_m = 1.25$ m **(b)** $h_m = 1.5$ m

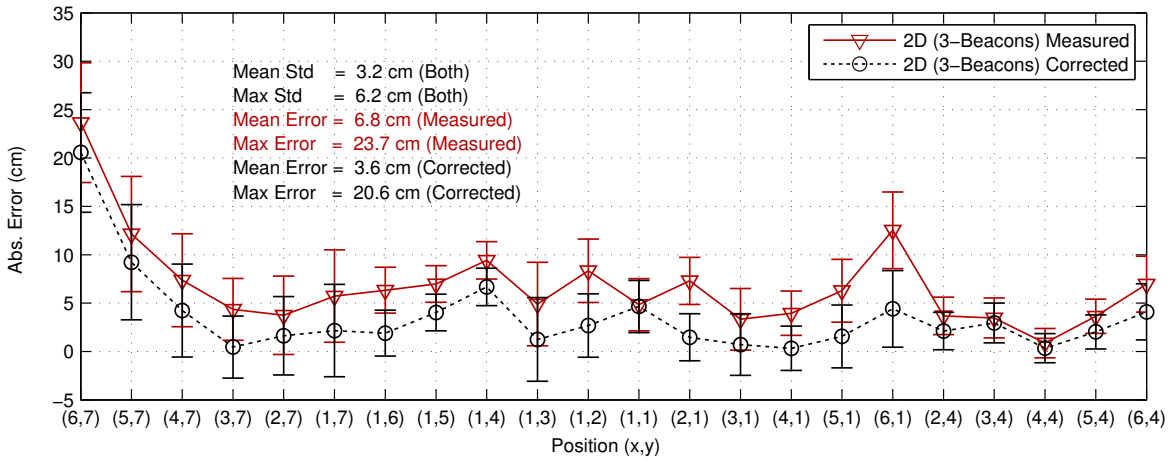


(c) $h_m = 1.75$ m **(d)** $h_m = 2$ m

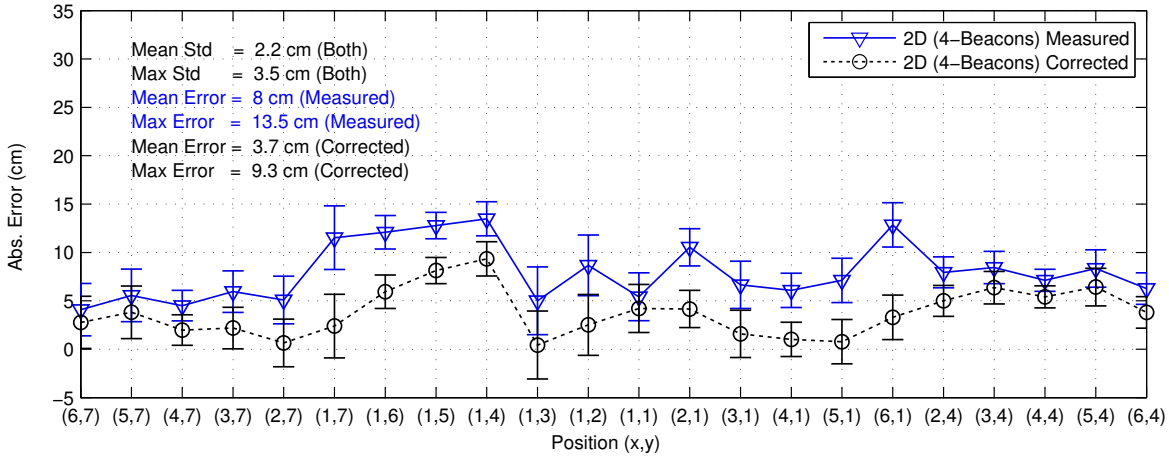
Figure 6.6: Absolute 2D error when the height of the microphone in the target node changes between 1.25 m and 2 m for the room used in the previous experiments for three beacons.

6.3 Akkurate Positioning System Evaluation

Figures 6.7a and 6.7b show the statistical results for the measured data and its corrected version, i.e. when the error imposed by the non-coplanarity previously discussed is removed. In addition, a summary of the statistical metrics used to evaluate the 2D system performance is presented when three and four beacons are used. The evaluation is based in the absolute max/mean error plus standard deviation and in the absolute max error obtained for 95% of the cases. Very stable and accurate position estimates were obtained with a mean absolute standard deviation of 2.2 cm (4-Beacons) and 3.2 cm (3-Beacons). In addition to this, an absolute positioning error of less than 10 cm and 16 cm in 95% of the cases was achieved for four and three beacons respectively, thus meeting the requirements of the applications we are targeting.



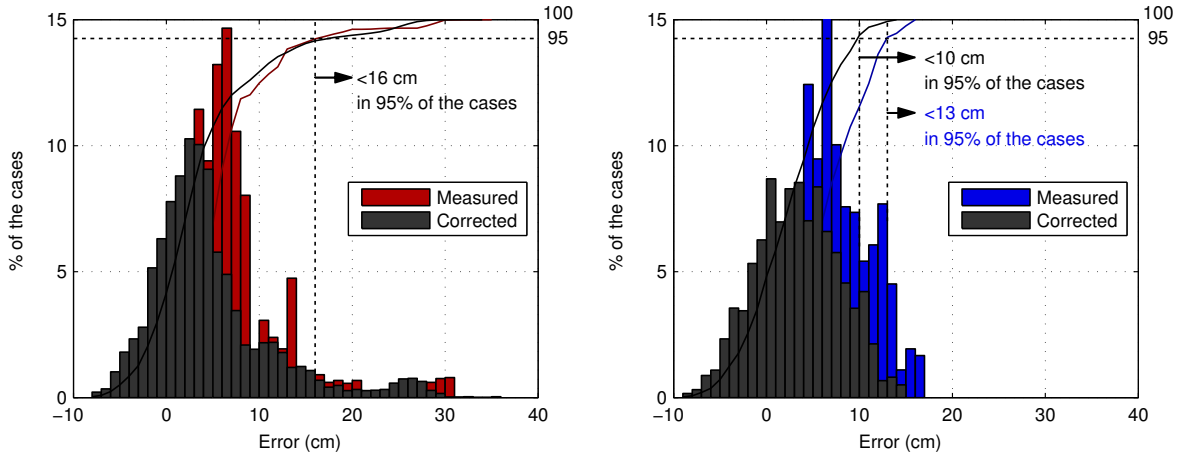
(a) 2D statistical results using 3-Beacons.



(b) 2D statistical results using 4-Beacons.

Figure 6.7: Experiment A. Statistical results. Measured and corrected absolute mean error and standard deviation, for each position evaluated, when three and four beacons are used.

Figures 6.8a and 6.8b present the error histogram and the cumulative sum function results for each for the data presented in Figures 6.7a and 6.7b, respectively.



(a) 3-Beacons case.

(b) 4-Beacons case.

Figure 6.8: Error histograms and cumulative sum plots for each case.

6.4 Final Discussion

To compare the proposed system with other positioning systems, we will re-use Table 2.2 from Section 2.2.13. Table 6.1 is a replica of Table 2.2 with the results of our system added. From the new table it is possible to observe that most of the selected acoustic systems have centimeter-level accuracy and all of them use time-based methods in the ranging process thus reaching generally the best performance in terms of accuracy among all the presented systems. In relation to the smartphone compatibility feature, only the Guoguo and the Beep systems are eligible.

Furthermore, RF-based systems based in WiFi/FP showed to have accuracies below 2 m in 70% of the cases (COMPASS system) and below 4,7 m in 75% of the cases (RADAR system), which is far from acceptable for the application we are focused. On the other hand, the UWB approaches (UBISENSE, Pietrzyk et al. and Segura et al.) generally present centimeter-level accuracy but all of them rely in complex and expensive hardware that is not compatible with conventional smartphones. The Nokia HAIP and Ubisense systems are both used in applications were increased accuracy and high position rate of estimations is need. Nokia HAIP [37] uses a dedicated and expensive infrastructure of "Locators" (antenna arrays) that are used to measure the Angle of Arrival (AoA) of a Bluetooth packet that is continuously transmitted by the smartphone. A major drawback of these two systems is the high cost of the infrastructure.

The main requirements imposed in the system design (e.g. Smartphone Compatibility, Centimeter-level Accuracy, High Doppler Tolerance and Low-Cost Infrastructure) resulted in design decisions that proved to be appropriate for the target application. The accuracy obtained meets the specified

criteria for users moving at higher speeds, and the use of non-invasive audio pulses overcomes the problems presented by the audible signals used in [36], [32] and [31].

From the Table 6.1 it is possible to observe an additional smartphone compatible system [30]. This system uses "unnoticeable to humans" audio/US pulses for ranging. They use a basic synchronized mechanism similar to [33, 35] with RF tokens triggered by a central node. Although the authors present a preliminary study based only on acoustic background noise measurements with the system working, they did not evaluate the signals with subjects; therefore, their classification of "unnoticeable to humans" signals is unfounded.

A previous work by ourselves [25] was added to Table 6.1 for comparison. In this work we reported a 2D absolute mean error of 9.6 cm and a mean standard deviation of 0.8 cm. This increase in precision performance can be easily observed by the resulting standard deviation value of 0.8 cm, which is significantly smaller than the results we are reporting in this thesis. The performance reported in [25] shows a similar accuracy but a higher precision that can be justified by these two main factors:

- The system uses ToA measurements for ranging (all beacons and mobile station are in sync);
- The system uses broadband ultrasonic pulses with large TBP that resulted in narrow compressed peaks and thus in more precise range measurements.

When compared to other centimeter-level positioning systems, the system proposed in this thesis stands out for several reasons, among which we highlight:

- Smartphone compatibility, i.e. users do not need any special hardware due to the use of MT-TDoA ranging which results in synchronization-free mobile devices. The limited bandwidth available for the acoustic signal design was mitigated by the use of an improved MT-TDoA estimation approach;
- Non-invasive pulse design enables effective use of the available bandwidth. The proposed approaches for non-invasive pulse design resulted in a reduced perceptual impact tested experimentally with subjects;
- The proposed Automatic Time Synchronization and Syntonization (ATSS) protocol takes advantage of the already existent RF transceiver built into the beacon for both communications and synchronization. Main advantages of this approach include: simple hardware design (less components used); cost reduction (uses only one radio); reduced RF interference (less bands used) and security increase (sync data can be encrypted);
- Low-cost infrastructure (approximately 35€ per beacon) due to the commercial-of-the-shelf component selection in the acoustic beacon design;
- Self-positioning (i.e. GPS-like) was possible due to the inclusion of a synchronization protocol shared by the beacon infrastructure. Additionally, due to the low interference that passive receiving ranging presents when compared to the active transmitting mode, it is possible to use the system with a high number of simultaneous users (unlimited number theoretically).

Name	Category	Year	Technology	Ranging Method	Reported Accuracy	Simultaneous Users	Bandwidth	Update Rate	Cost	Smartphone Compatibility
Akuratte	**Acoustic**	**2014**	**audio/US**	**MT-TDoA**	**<10 cm** (95%)	**high**	**18-22 kHz**	**$\leq$ 3 Hz**	**Inexpensive**	**yes**
Guoguo [30]	Acoustic	2013	Audio/US	MT-TDoA	6-25 cm	High	17-20 kHz	$\leq$ 2 Hz	Inexpensive	yes
Lopes et al. [25]	**Acoustic**	**2012**	**US**	**ToA**	**9.6 cm**	**high**	**20-45 kHz**	**$\leq$ 5 Hz**	**Inexpensive**	**no**
LOSNUS [31]	Acoustic	2010	Audio/US	MT-TDoA	1 cm	High	35-65 kHz	$\leq$ 10 Hz	Expensive	no
3D Locus [32]	Acoustic	2009	Audio/US	RToF	0.5 cm (90%)	Limited	<25 kHz	10 Hz	Inexpensive	no
Parrot [33]	Acoustic	2006	US+auxRF	ST-TDoA	2 cm ($\pm$ 4)	High	1 kHz	$\leq$5 Hz	Expensive	no
Cricket [35]	Acoustic	2005	US+auxRF	ST-TDoA	1-2 cm	High	1 kHz	1 Hz	Inexpensive	no
Beep [36]	Acoustic	2005	Audio	ToA	61 cm(97%)	Limited	4 kHz	-	Inexpensive	yes
Nokia HAIP [37]	RF	2012	Bluetooth	AoA	30-100 cm	High	1 MHz	0.1-40 Hz	Expensive	yes
UBISENSE [38]	RF	2011	UWB	AoA/MT-TDoA	<15 cm	Limited	-	-	Expensive	no
Pietrzyk et al. [42]	RF	2010	UWB	ToA	1-2 cm	Limited	-	-	Expensive	no
Segura et al. [43]	RF	2010	UWB	MT-TDoA	<20 cm	Limited	-	-	Expensive	no
COMPASS [40]	RF	2005	WiFi/FP	RSS	<2 m (70%)	High	-	-	Inexpensive	yes
RADAR [39]	RF	2000	WiFi/FP	RSS	<4.7 m (75%)	High	-	0.25 Hz	Inexpensive	yes

Table 6.1: Performance of selected centimetre-level positioning systems presented in Table 2.2 with the works by ourselves added.

CHAPTER 7

Conclusions

Accurate head positioning is important for indoor applications, such as augmented reality, virtual reality, gaming, and audio guiding applications. This work explored the possibility of estimating the head position (with centimeter-level accuracy) of a user carrying a smartphone in a room, without any special hardware to operate. The main goal was to develop a multi-user and high accuracy head tracking system fully compatible with conventional smartphones. The system was designed to be the basis of an augmented reality audio guiding system for museums.

Smartphone-based centimeter-level positioning is an open problem, which can be confirmed by the lack of commercial and affordable solutions, cf. Chapter 2. To achieve centimeter-level positioning accuracy in a smartphone-based system, Time-of-Flight (ToF) ranging was used with non-invasive audio pulses specially designed to reduce the perceptual impact on humans. The existing general theoretical literature on the subject, and specifically on loudness perception for the band of interest, was inconclusive, and this lead us to perform additional research.

Next, we recall the research questions presented in Section 1.1.2:

- Is it possible to use the audio band to perform effective non-invasive acoustic ranging? How far is it possible to reduce perceptual impact on humans?

- Is it possible to effectively estimate the position of a conventional smartphone (without any special hardware) based in non-invasive acoustic ranging?

- Since lower frequency content will be used, multipath and interference will increase considerably due to the lower attenuation that signals experience. Is it possible to effectively mitigate these problems to increase the performance of the positioning system?

The first research question is answered by the work presented in Sections 3.5 and 3.6. The perceptual evaluation results, reported in Section 6.1 show that it is possible to use the audible band for ranging without annoying users. The proposed approaches for non-invasive pulse design in the band 16-24 kHz were evaluated with subjects and the results achieved are in close agreement with the theory behind loudness perception for complex stimuli with durations below 200 ms. The data observed in Figures 6.2 and 6.3 indicates that the stimuli with lower durations tend to be less audible.

Moreover, if we look at the bandwidth occupied by the stimuli, the BPSK-based pulses occupy all the available bandwidth (16-24 kHz) with invariable spectral power density. On the other hand, the chirp-based pulses have a time varying spectral power occupation. This leads us to the evidence that the BPSK-based pulses tend to be more perceptible due to their invariable bandwidth occupation. This evidence can be justified by the spectral loudness summation increase that BPSK-based pulses experience, when compared to the chirp-based pulses, i.e. more critical bands are occupied by the BPSK-based stimuli which reinforces the loudness sensation.

The proposed definition for non-invasive audio was generally accepted by the subjects. Even when they were able to detect the presence of the stimulus, all agreed that general classification of the stimulus as non-invasive audio was fairly acceptable.

The second and third research questions obtained an answer based on the work described on Chapters 4 and 5, and in the results described on Sections 6.2 and 6.3.

The first finding is related to development of an Automatic Time Synchronization and Syntonization (ATSS) protocol (Chapter 4), that ensures a sync offset mean error below 0.86 μs (sync period of 5 s). This enabled us to keep the WSN infrastructure in sync and use MT-TDoA ranging in the measurement stage. Experimental tests have shown that the ATSS protocol imposes an offset error with a standard deviation below 1.25 μ s (sync period of 1 s). This resulted in a standard deviation of less than 9 mm in distance measurements (validated for distances up to 8 m) between two acoustic motes.

Secondly, in Chapter 5, is introduced an improved MT-TDoA estimation approach (that takes advantage of the beacon signals' periodicity) and by performing Non-Line-of-Sight (NLoS) mitigation, we were able to obtain very stable and accurate position estimates with an absolute error of less than 10 cm in 95% of the cases, and a mean absolute standard deviation of 2.2 cm for a position refresh period of 350 ms. These results are shown at the end of Chapter 5 and were obtained considering static positions in a standard room with dimensions $9 \times 8 \times 3$ m.

Additionally, a qualitative evaluation was also performed when a moving person equipped with a receiver on top of his/her head moved in two trajectories with different speeds, the first moving at approximately 0.5 m/s and the second moving at approximately 1.5 m/s. The use of a chirp-based pulse in the experimental evaluation was important for the performance obtained for both trajectories, due to its Doppler resilience (Section 3.5.3), when combined with a matched filter for pulse detection.

7.1 Future Work

To improve the performance of the proposed system, when severe interference and NLoS situations may occur, a possible direction is to develop a navigation filter that combines the output of the proposed indoor positioning system with the data coming from the Inertial Measurement Unit (IMU) natively available in the smartphone.

This will allow to use the information given by the accelerometer, gyroscope and magnetometer inside the device to obtain heading information, and its related dynamics. In [109] we have presented a preliminary study with an iPhone 4s for heading estimation using the Core Location Framework to obtain the heading of the smartphone when it is in rotation around one axis for more than 5 minutes. Without the magnetometer the error drifts to approximately 100° (approximately 20° per minute). By performing sensor fusion with the information from the magnetometer the drift error reduces to approximately 2°, which is acceptable for our application case.

Another considerable improvement can come from the introduction of a wireless (e.g. Bluetooth) external IMU with magnetometer, placed at the user head. This can provide heading information (head orientation), and inertial information about the user dynamics. This way, we can include a simplified short-time inertial navigation model and a proper data fusion algorithm to estimate the user position based on his/her dynamics for small periods when the positioning information is corrupted by interference, or inaccurate due to NLoS.

Additionally, an interference detection and classification method can be introduced in the system to discard invalid sets of MT-TDoA measurements that have passed the existing validation mechanism.

References

[1] Y. Liu, M. Storring, T. Moeslund, C. Madsen, and E. Granum, "Computer vision based head tracking from re-configurable 2D markers for AR," in *Mixed and Augmented Reality, 2003. Proceedings. The Second IEEE and ACM International Symposium on*, pp. 264–267, 2003.

[2] J. Jun, Q. Yue, and Z. Qing, "An extended marker-based tracking system for augmented reality," in *Modeling, Simulation and Visualization Methods (WMSVM), 2010 Second International Conference on*, pp. 94–97, 2010.

[3] E. Murphy-Chutorian and M. Trivedi, "Head pose estimation and augmented reality tracking: An integrated system and evaluation for monitoring driver awareness," *Intelligent Transportation Systems, IEEE Transactions on*, vol. 11, no. 2, pp. 300–311, 2010.

[4] C. Wong, Z. Zhang, S. McKeague, and G.-Z. Yang, "Multi-person vision-based head detector for markerless human motion capture," in *Body Sensor Networks (BSN), 2013 IEEE International Conference on*, pp. 1–6, 2013.

[5] X. Mou and H. Wang, "A fast and robust head pose estimation system based on depth data," in *Robotics and Biomimetics (ROBIO), 2012 IEEE International Conference on*, pp. 470–475, 2012.

[6] X. Yu, W. Chua, L. Dong, K. E. Hoe, and L. Li, "Head pose estimation in thermal images for human and robot interaction," in *Industrial Mechatronics and Automation (ICIMA), 2010 2nd International Conference on*, vol. 2, pp. 698–701, 2010.

[7] J. Foytik, V. Asari, M. Youssef, and R. Tompkins, "Head pose estimation from images using canonical correlation analysis," in *Applied Imagery Pattern Recognition Workshop (AIPR), 2010 IEEE 39th*, pp. 1–7, 2010.

[8] R. Mautz, *Indoor Positioning Technologies*. Institute of Geodesy and Photogrammetry, Department of Civil, Environmental and Geomatic Engineering, ETH Zurich, 2012.

[9] S. A. R. Zekavat and R. M. Buehrer, *Handbook of Position Location - Theory, Practice, and Advances*. John Wiley & Sons, Inc., 2011.

[10] Y. Suzuki and H. Takeshima, "Equal-loudness-level contours for pure tones.," *The Journal of the Acoustical Society of America*, vol. 116, no. 2, pp. 918–933, 2004.

[11] R. Watson, G. Lachapelle, R. Klukas, S. Turunen, S. Pietilä, and I. Halivaara, "Investigating GPS Signals Indoors with Extreme High-Sensitivity Detection Techniques," *Journal of The Institute of Navigation,*, vol. 52, no. 4, pp. 199–214, Winter 2005-2006.

[12] G. Dedes and A. G. Dempster, "Indoor GPS Positioning," in *Proceedings of the IEEE Semi-annual Vehicular Technology Conference*, Citeseer, 2005.

[13] Vicon Motion Systems Ltd. UK, "Vicon tracking system." `http://www.vicon.com`, [Online; accessed 10-July-2014].

[14] Y.-R. Chen, C.-M. Huang, and L.-C. Fu, "Visual tracking of human head and arms with a single camera," in *Intelligent Robots and Systems (IROS), 2010 IEEE/RSJ International Conference on*, pp. 3416–3421, 2010.

[15] F. Kondori, S. Yousefi, H. Li, S. Sonning, and S. Sonning, "3D head pose estimation using the Kinect," in *Wireless Communications and Signal Processing (WCSP), 2011 International Conference on*, pp. 1–4, 2011.

[16] J. Dai, "Head Pose Estimation Using Kinect." `http://www.jwdai.net/Research/HeadPoseKin/HeadPoseKin.htm`, 2012, [Online; accessed 14-July-2014].

[17] Fakespace Corporation, "Miscellaneous Stereo 3D Hardware - Boom mounted Displays." `http://www.stereo3d.com/more.htm`, 2006, [Online; accessed 10-July-2014].

[18] M. Bolas, "Human factors in the design of an immersive display," *Computer Graphics and Applications, IEEE*, vol. 14, no. 1, pp. 55–59, 1994.

[19] InterSense, "InertiaCube BT^{TM}." `http://www.intersense.com/pages/18/60/`. [Online; accessed 19-July-2014].

[20] L. Porzi, E. Ricci, T. Ciarfuglia, and M. Zanin, "Visual-inertial tracking on Android for Augmented Reality applications," in *Environmental Energy and Structural Monitoring Systems (EESMS), 2012 IEEE Workshop on*, pp. 35–41, 2012.

[21] U. Schnabel, M. Hegenloh, H. Müller, and M. Zehetleitner, "Electromagnetic tracking of motion in the proximity of computer generated graphical stimuli: A tutorial.," *Behavior research methods*, vol. 45, pp. 696–701, 09 2013.

[22] A. Markham and N. Trigoni, "Underground Localization in 3-D Using Magneto-Inductive Tracking," *IEEE Sensors Journal*, vol. 12, pp. 1809–1816, June 2012.

[23] Polhemus Corporation, "Polhemus Head Tracker." `http://www.polhemus.com/motion-tracking/head-trackers/`, [Online; accessed 12-July-2014].

[24] J. N. Vieira, S. I. Lopes, C. A. C. Bastos, and P. N. Fonseca, "Ultrasound sensor array for robust location," in *Multi-Agent Robotic Systems, Proceedings* (Sapaty, P and Filipe, J, ed.), pp. 84–93, 2007. 3rd International Workshop on Multi-Agent Robotic Systems, Angers, FRANCE, MAY, 2007.

[25] S. I. Lopes, J. M. N. Vieira, and D. Albuquerque, "High Accuracy 3D Indoor Positioning Using Broadband Ultrasonic Signals," *2012 11th IEEE International Conference on Trust, Security and Privacy in Computing and Communications*, vol. 0, pp. 2008–2014, 2012.

[26] Y. Cheng, X. Wang, T. Caelli, and B. Moran, "Target tracking and localization with ambiguous phase measurements of sensor networks," in *Acoustics, Speech and Signal Processing (ICASSP), 2011 IEEE International Conference on*, pp. 4048–4051, 2011.

[27] N. Patwari, J. N. Ash, S. Kyperountas, A. O. Hero, R. L. Moses, and N. S. Correal, "Locating the nodes: cooperative localization in wireless sensor networks," *Signal Processing Magazine, IEEE*, vol. 22, no. 4, pp. 54–69, 2005.

[28] S. Gezici, Z. Tian, G. B. Giannakis, H. Kobayashi, A. F. Molisch, H. V. Poor, and Z. Sahinoglu, "Localization via Ultra-Wideband Radios," *IEEE Signal Processing Magazine*, vol. 22, pp. 70–84, July 2005.

[29] B. Dil, S. Dulman, and P. Havinga, "Range-Based Localization in Mobile Sensor Networks," in *Wireless Sensor Networks* (K. Römer, H. Karl, and F. Mattern, eds.), vol. 3868 of *Lecture Notes in Computer Science*, pp. 164–179, Springer Berlin / Heidelberg, 2006.

[30] K. Liu, X. Liu, , and X. Li, "Guoguo: Enabling Fine-grained Indoor Localization via Smartphone," in *MobiSys'13*, Taipei, Taiwan, June 25-28, 2013.

[31] H. Schweinzer and M. Syafrudin, "LOSNUS: An ultrasonic system enabling high accuracy and secure TDoA locating of numerous devices," in *Indoor Positioning and Indoor Navigation (IPIN), 2010 International Conference on*, pp. 1 –8, sept. 2010.

[32] J. Prieto, A. Jimenez, J. Guevara, J. Ealo, F. Seco, J. Roa, and F. Ramos, "Performance Evaluation of 3D-LOCUS Advanced Acoustic LPS," *Instrumentation and Measurement, IEEE Transactions on*, vol. 58, pp. 2385–2395, Aug 2009.

[33] W. Zhang, J. Djugash, and S. Singh, "Parrots: A range measuring sensor network," tech. rep., Robotics Institute, Carnegie Mellon University, Pittsburgh, Pennsylvania 15213, 2006.

[34] N. B. Priyantha, A. Chakraborty, and H. Balakrishnan, "The Cricket location-support system," in *The 6th annual international conference on Mobile computing and networking*, (Boston, Massachusetts, United States), pp. 32–43, ACM Press, 2000.

[35] N. B. Priyantha, *The Cricket Indoor Location System*. PhD thesis, Massachusetts Institute of Technology, 2005.

[36] A. Mandal, C. Lopes, T. Givargis, A. Haghighat, R. Jurdak, and P. Baldi, "Beep: 3D Indoor Positioning using Audible Sound," in *Consumer Communications and Networking Conference, 2005. CCNC. 2005 Second IEEE*, pp. 348–353, Jan 2005.

[37] J. Rantala, "High Accuracy Indoor Positioning - Technology Solution and Business Implications," in *3rd Invitational Workshop on Opportunistic RF Localization for Next Generation Wireless Devices*, NOKIA Research Center, 2012.

[38] Ubisense Corporation, *Ubisense Series 7000 IP Sensors, Precision ultra-wideband measurement devices for industrial environments*. Ubisense, 2012.

[39] P. Bahl and V. N. Padmanabhan, "RADAR: an in-building RF-based user location and tracking system," in *Nineteenth Annual Joint Conference of the IEEE Computer and Communications Societies*, vol. 2, pp. 775–784 vol.2, 2000.

[40] F. Kargl and A. Bernauer, "The COMPASS location system," in *Proceedings of the First international conference on Location- and Context-Awareness*, LoCA'05, (Berlin, Heidelberg), pp. 105–112, Springer-Verlag, 2005.

[41] M. I. T. Computer Science and Artificial Intelligence Laboratory, "The Cricket Indoor Location System." `http://cricket.csail.mit.edu/`, 2003, [Online; accessed 10-July-2014].

[42] M. Pietrzyk and T. von der Grun, "Experimental validation of a TOA UWB ranging platform with the energy detection receiver," in *Indoor Positioning and Indoor Navigation (IPIN), 2010 International Conference on*, pp. 1–8, Sept 2010.

[43] M. Segura, H. Hashemi, C. Sisterna, and V. Mut, "Experimental demonstration of self-localized Ultra Wideband indoor mobile robot navigation system," in *Indoor Positioning and Indoor Navigation (IPIN), 2010 International Conference on*, pp. 1–9, Sept 2010.

[44] I. Amundson and X. D. Koutsoukos, *A Survey on Localization for Mobile Wireless Sensor Networks*. Springer-Verlag Berlin Heidelberg, 2009.

[45] L. M. P. L. de Brito and L. M. R. Peralta, "An Analysis of Localization Problems and Solutions in Wireless Sensor Networks," *Revista de Estudos Politécnicos*, vol. VI, no. 9, 2008.

[46] A. Srinivasan and J. Wu, "A Survey on Secure Localization in Wireless Sensor Networks," in *Encyclopedia of Wireless and Mobile Communications*, 2008.

[47] H. Liu, H. Darabi, P. Banerjee, and J. Liu, "Survey of Wireless Indoor Positioning Techniques and Systems," *Systems, Man, and Cybernetics, Part C: Applications and Reviews, IEEE Transactions on*, vol. 37, pp. 1067 –1080, nov. 2007.

[48] A. Koubaa and M. B. Jamaa, "Taxonomy of Fundamental Concepts of Localization in Cyber-Physical and Sensor Networks," tech. rep., CISTER Research Unit, Polytechnic Institute of Porto (ISEP-IPP), February 2013.

[49] N. Patwari and A. O. Hero, III, "Demonstrating distributed signal strength location estimation," in *SenSys '06: Proceedings of the 4th international conference on Embedded networked sensor systems*, (New York, NY, USA), pp. 353–354, ACM, 2006.

[50] R. Masiero, *RSSI based Tracking Algorithms for Wireless Sensor Networks: Theoretical aspects and Performance Evaluation*. PhD thesis, Universitá degli studi di Padova, 2007.

[51] P. Claro and N. B. Carvalho, "Local Positioning System Based on Artificial Neural Networks," *ICANN*, pp. 699–708, Porto 2007.

[52] K. Ho and Y. Chan, "Solution and performance analysis of geolocation by TDOA," *Aerospace and Electronic Systems, IEEE Transactions on*, vol. 29, pp. 1311–1322, Oct 1993.

[53] A. Youssef and M. Youssef, "A Taxonomy of Localization Schemes for Wireless Sensor Networks," in *The 2007 International Conference on Wireless Networks (ICWN'07),Las Vegas, Nevada, June 25-28, 2007*, Juni 2007.

[54] D. Niculescu and B. Nath, "Ad hoc positioning system (APS) using AOA," in *INFOCOM 2003. Twenty-Second Annual Joint Conference of the IEEE Computer and Communications. IEEE Societies*, vol. 3, pp. 1734 – 1743 vol.3, 2003.

[55] L.-W. Yeh, M.-S. Hsu, Y.-F. Lee, and Y.-C. Tseng, "Indoor localization: Automatically constructing today's radio map by iRobot and RFIDs," in *Sensors, 2009 IEEE*, pp. 1463–1466, Oct 2009.

[56] R. Hansen, R. Wind, C. S. Jensen, and B. Thomsen, "Algorithmic strategies for adapting to environmental changes in 802.11 location fingerprinting," in *Indoor Positioning and Indoor Navigation (IPIN), 2010 International Conference on*, pp. 1–10, Sept 2010.

[57] A. H. Sayeda, A. Tarighata, and N. Khajehnouri, "Network-based wireless location," *IEEE Signal Processing Magazine*, vol. 22, no. 4, pp. 24–40, 2005.

[58] P. Bergamo, S. Asgari, H. Wang, D. Maniezzo, L. Yip, R. E. Hudson, K. Yao, and D. Estrin, "Collaborative Sensor Networking Towards Real-Time Acoustical Beamforming in Free-Space and Limited Reverberance," *IEEE Transactions on Mobile Computing*, vol. 3, no. 3, pp. 211–224, 2004.

[59] J. M. N. Vieira, S. I. Lopes, C. A. C. Bastos, and P. N. Fonseca, "Sistema de localização utilizando ultra-sons," *Encontro Científico do Festival Nacional de Robótica 2006. Guimarães, Portugal*, April 2006.

[60] R. Stoleru, T. He, J. A. Stankovic, and D. Luebke, "A high-accuracy, low-cost localization system for wireless sensor networks," in *SenSys '05: Proceedings of the 3rd international conference on Embedded networked sensor systems*, (New York, NY, USA), pp. 13–26, ACM, 2005.

[61] J. Zhang, T. Yan, J. A. Stankovi, and S. H. Son, "Thunder: towards practical, zero cost acoustic localization for outdoor wireless sensor networks," *SIGMOBILE Mob. Comput. Commun. Rev.*, vol. 11, no. 1, pp. 15–28, 2007.

[62] A. Savvides, M. Srivastava, L. Girod, and D. Estrin, *Localization in sensor networks*. Norwell, MA, USA: Kluwer Academic Publishers, 2004.

[63] G. Mao, B. Fidan, G. Mao, and B. Fidan, *Localization Algorithms and Strategies for Wireless Sensor Networks*. Hershey, PA: Information Science Reference - Imprint of: IGI Publishing, 2009.

[64] D. Self, R. Brice, B. Duncan, J. Linsley, H. I. Sinclair, A. Singmin, D. Davis, E. Patronis, and J. Watkinson, *Audio Engineering - know it all*. Elsevier, 2009.

[65] D. M. Howard and J. A. S. Angus, *Acoustics and Psychoacoustics*. Focal Press, Elsevier, fourth ed., 2009.

[66] L. E. Kinsler, A. R. Frey, A. B. Coppens, and J. V. Sanders, *Fundamentals Of Acoustics*. Wiley, 4th ed., 1999.

[67] H. E. Bass, L. C. Sutherland, A. J. Zuckerwar, D. T. Blackstock, and D. M. Hester, "Atmospheric absorption of sound: Further developments," *The Journal of the Acoustical Society of America*, vol. 97, pp. 680–683, 1995.

[68] H. E. Bass, L. C. Sutherland, A. J. Zuckerwar, D. T. Blackstock, and D. M. Hester, "Erratum: Atmospheric absorption of sound: Further developments [J. Acoust. Soc. Am. 97, 680–683 (1995)]," *The Journal of the Acoustical Society of America*, vol. 99, pp. 1259–1259, 1996.

[69] H. E. Bass, L. C. Sutherland, and Zuckerwar, "Atmospheric absorption of sound: Update," *The Journal of the Acoustical Society of America*, vol. 88, pp. 2019–2021, 1990.

[70] C. M. Harris, *Dictionary of Architecture and Construction*. McGraw-Hill Companies, Inc., 2003.

[71] *Environmental Noise*. Bruel & Kjaer Sound & Vibration Measurement, Copyright © 2000, 2001.

[72] D. Albuquerque, J. M. N. Vieira, S. I. Lopes, C. A. C. Bastos, and P. J. S. G. Ferreira, "Indoor Acoustic Simulator for Ultrasonic Broadband Signals with Doppler Effect," *Journal of Applied Acoustics - Elsevier*, Submitted in September 2014.

[73] D. Havelock, S. Kuwano, and M. Vorländer, eds., *Handbook of Signal Processing in Acoustics*. Springer US, 2008.

[74] H. J. Foley and M. W. Matlin, "Sensation & Perception. The Auditory System." `http://www.skidmore.edu/~hfoley/Perc9.htm`, [Online; accessed 5-February-2014].

[75] S. D. Smith, C. W. Nixon, and H. E. V. Gierke, *Noise and Vibration Control Engineering: Principles and Applications*. John Wiley and Sons New York, 2006.

[76] C. Q. Howard, C. H. Hansen, and A. C. Zander, "A Review of Current Ultrasound Exposure Limits," *Journal of Occupational Health and Safety - Australia and New Zealand*, vol. 21, no. 3, pp. 253–257, 2005.

[77] M. Florentine, "Loudness," in *Loudness* (M. Florentine, A. N. Popper, and R. R. Fay, eds.), vol. 37 of *Springer Handbook of Auditory Research*, pp. 1–15, Springer New York, 2011.

[78] H. Fletcher and W. A. Munson, "Loudness, Its Definition, Measurement and Calculation," *Bell System Technical Journal*, vol. 12, pp. 377–430, 1933.

[79] S. S. Stevens, "A scale for the measurement of a psychological magnitude: loudness.," *Psychological Review*, vol. 43, pp. 405–416, Sept. 1936.

[80] D. W. Robinson and R. S. Dadson, "A re-determination of the equal-loudness relations for pure tones," *British Journal of Applied Physics*, vol. 7, no. 5, p. 166, 1956.

[81] H. Fastl and E. Zwicker, *Psychoacoustics: Facts and Models (Springer Series in Information Sciences)*. Springer, 3rd ed., 2006.

[82] *ISO 226:2003 - Acoustics: Normal equal-loudness-level contours*. International Standard Organization, 2003.

[83] C. B. Pedersen and G. Salomon, "Temporal integration of acoustic energy," *Acta Oto-laryngologica*, vol. 83, no. 1-6, pp. 417–423, 1977.

[84] D. Algom, A. Rubin, and L. Cohen-Raz, "Binaural and temporal integration of the loudness of tones and noises," *Perception & Psychophysics*, vol. 46, no. 2, pp. 155–166, 1989.

[85] J. O. Smith, *Introduction to Digital Filters with Audio Applications*. W3K Publishing, 2007.

[86] S. M. Kay, *Fundamentals of Statistical Signal Processing, Volume 2: Detection Theory*. Prentice Hall, 1993.

[87] N. Levanon and E. Mozeson, *Radar Signals*. Hoboken, New Jersey: John Wiley & Sons, Inc., 2004.

[88] N. Z. Søren Bech, *Perceptual Audio Evaluation - Theory, Method and Application*. John Wiley & Sons, Inc., 2006.

[89] S. W. Golomb and G. Gong, *Signal Design for Good Correlation for Wireless Communication, Cryptography, and Radar*. Cambridge University Press, 2005.

[90] B. Sundararaman, U. Buy, and A. D. Kshemkalyani, "Clock synchronization for wireless sensor networks: a survey," *Ad Hoc Networks*, vol. 3, no. 3, pp. 281–323, 2005.

[91] E. Serpedin and Q. M. Chaudhari, *Synchronization in Wireless Sensor Networks: Parameter Estimation, Peformance Benchmarks, and Protocols*. New York, NY, USA: Cambridge University Press, 1st ed., 2009.

[92] "IEEE Standard for a Precision Clock Synchronization Protocol for Networked Measurement and Control Systems," *IEEE Std 1588-2008 (Revision of IEEE Std 1588-2002)*, pp. c1 –269, 24 2008.

[93] H. Cho, J. Jung, B. Cho, Y. Jin, S.-W. Lee, and Y. Baek, "Precision Time Synchronization using IEEE 1588 for Wireless Sensor Networks," in *Computational Science and Engineering, 2009. CSE '09. International Conference on*, vol. 2, pp. 579 –586, aug. 2009.

[94] D. Wobschall and Y. Ma, "Synchronization of Wireless Sensor Networks using a modified IEEE 1588 protocol," in *Precision Clock Synchronization for Measurement Control and Communication (ISPCS), 2010 International IEEE Symposium on*, pp. 67 –70, 27 2010-oct. 1 2010.

[95] Texas Intruments, "SimpliciTI Compliant Protocol Stack." `http://www.ti.com/tool/SIMPLICITI`, [Online; accessed 20-July-2014].

[96] F. Gustafsson and F. Gunnarsson, "Mobile positioning using wireless networks: possibilities and fundamental limitations based on available wireless network measurements," *Signal Processing Magazine, IEEE*, vol. 22, pp. 41–53, July 2005.

[97] S. M. Kay, *Fundamentals of Statistical Signal Processing, Volume 1: Estimation Theory*. Prentice Hall, 1993.

[98] F. Gustafsson and F. Gunnarsson, "Positioning using time-difference of arrival measurements," in *Acoustics, Speech, and Signal Processing, 2003. Proceedings. (ICASSP '03). 2003 IEEE International Conference on*, vol. 6, pp. VI–553–6 vol.6, April 2003.

[99] J. Chaffee and J. Abel, "GDOP and the Cramer-Rao bound," in *Position Location and Navigation Symposium, 1994., IEEE*, pp. 663–668, Apr 1994.

[100] C. Mensing and S. Plass, "Positioning Algorithms for Cellular Networks using TDOA," in *Acoustics, Speech and Signal Processing, 2006. ICASSP 2006 Proceedings. 2006 IEEE International Conference on*, vol. 4, pp. IV–IV, May 2006.

[101] A. Rolando and E. Amoruso, "An ubiquitous positioning system based on IEEE 802.15.4 radio signals," in *Indoor Positioning and Indoor Navigation (IPIN), 2013 International Conference on*, pp. 1–10, Oct 2013.

[102] R. B. Langley, "Dilution of Precision," *GPS World*, May 1999.

[103] J. L. Crassidis and J. L. Junkins, *Optimal Estimation of Dynamic Systems, Second Edition (Chapman & Hall/CRC Applied Mathematics & Nonlinear Science)*. Chapman & Hall/CRC, 2nd ed., 2011.

[104] J. McKay and M. Pachter, "Geometry optimization for GPS navigation," in *Decision and Control, 1997., Proceedings of the 36th IEEE Conference on*, vol. 5, pp. 4695–4699 vol.5, Dec 1997.

[105] Y. Chen, J.-A. Francisco, W. Trappe, and R. Martin, "A Practical Approach to Landmark Deployment for Indoor Localization," in *Sensor and Ad Hoc Communications and Networks, 2006. SECON '06. 2006 3rd Annual IEEE Communications Society on*, vol. 1, pp. 365 –373, sept. 2006.

[106] E. C. L. Chan and G. Baciu, *Introduction to Wireless Localization: With iPhone SDK Examples*. Wiley-IEEE Press, 1st ed., 2012.

[107] J. S. Bradley, "Using ISO 3382 measures, and their extensions, to evaluate acoustical conditions in concert halls," *Acoustical Science and Technology*, vol. 26, no. 2, pp. 170–178, 2005.

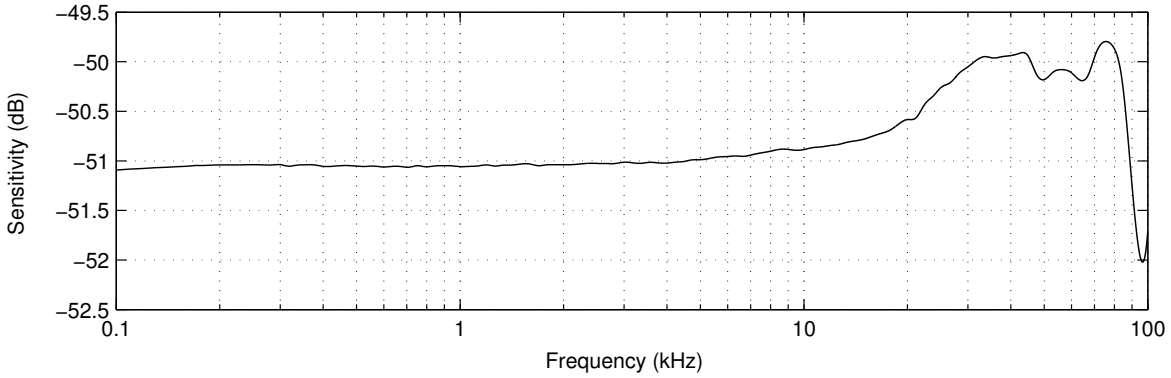
[108] J. Perez-Ramirez, D. Borah, and D. Voelz, "Optimal 3-D Landmark Placement for Vehicle Localization Using Heterogeneous Sensors," *Vehicular Technology, IEEE Transactions on*, vol. 62, pp. 2987–2999, Sept 2013.

[109] S. I. Lopes, J. M. N. Vieira, G. Campos, and P. Dias, "Sistema de Realidade Aumentada Áudio 3D para Dispositivos iOS," in *Proceedings of the 12th Audio Engineering Society Brazil Conference*, (São Paulo, Brazil), ISSN 2177-529X, 13-15 May 2014.

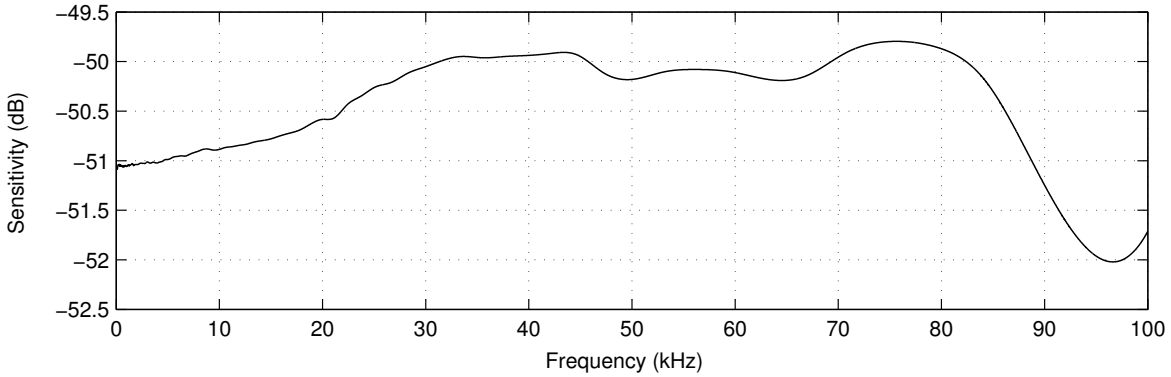
APPENDIX A

Brüel & Kjær 4954A reference microphone

The Brüel & Kjær 4954A is a laser welded 1/4 inch microphone capsule designed for high-level (Sensitivity of 2.8 mV/Pa and a dynamic range of 40 – 159 dB) and high-frequency (16 – 80000 Hz) measurements. The microphone frequency response was provided by the manufacturer in a Matlab file and is presented in Figure A.1.



(a) Logarithmic scale.



(b) Linear scale.

Figure A.1: Brüel & Kjær 4954A reference microphone frequency response.

APPENDIX B

Perceptual Evaluation Form

Name/Nome: _____ **Sex**/Sexo: ____ **Age**/Idade: ____ **Date**/Data:_____

Have you ever been diagnosed with hearing loss? Alguma vez lhe foi diagnosticado algum problema de audição? ____**Yes**/No ____**No**/Não.

Evaluation/Avaliação:

	Round 1 Ronda 1	**Round 2** Ronda 2	**Round 3** Ronda 3	**Round 4** Ronda 4
Audition 1 Audição 1				
Audition 2 Audição 2				
Audition 3 Audição 3				
Audition 4 Audição 4				
Audition 5 Audição 5				
Audition 6 Audição 6				
Audition 6 Audição 6				

Scale/Escala:

Qualitative Evaluation Avaliação Qualitativa	**Grade** Pontuação
Imperceptible Imperceptível	**1.0**
Perceptible, but not annoying Perceptível, mas não incomoda	**2.0**
Slightly annoying Incomoda um pouco	**3.0**
Annoying Incomoda	**4.0**
Very annoying Incomoda muito	**5.0**

APPENDIX C

Perceptual Evaluation Results

Subject 1 | **Sex**: M | **Age**: 27 | **Date**: 14/10/2014

	Round 1 55 dB (SPL)	Round 2 60 dB (SPL)	Round 3 65 dB (SPL)	Round 4 70 dB (SPL)	Round 5 75 dB (SPL)
Audition 1	1.0	1.0	1.1	1.0	1.1
Audition 2	1.1	1.1	1.0	2.2	2.5
Audition 3	1.1	1.1	2.7	3.1	3.0
Audition 4	1.0	1.0	1.5	2.6	2.5
Audition 5	1.1	1.0	1.0	2.3	2.3
Audition 6	1.2	1.2	1.1	2.5	3.5
Audition 7	1.0	1.0	1.0	1.0	1.0

Subject 2 | **Sex**: M | **Age**: 29 | **Date**: 14/10/2014

	Round 1 55 dB (SPL)	Round 2 60 dB (SPL)	Round 3 65 dB (SPL)	Round 4 70 dB (SPL)	Round 5 75 dB (SPL)
Audition 1	1.0	1.0	1.0	1.0	1.0
Audition 2	1.0	1.0	1.0	2.5	2.5
Audition 3	1.0	1.0	2.0	3.0	3.5
Audition 4	1.0	1.0	1.5	2.0	2.5
Audition 5	1.0	1.0	1.0	2.5	3.0
Audition 6	1.0	1.0	1.0	2.5	3.5
Audition 7	1.0	1.0	1.0	1.0	2.0

Subject 3 | **Sex**: M | **Age**: 30 | **Date**: 14/10/2014

	Round 1 55 dB (SPL)	Round 2 60 dB (SPL)	Round 3 65 dB (SPL)	Round 4 70 dB (SPL)	Round 5 75 dB (SPL)
Audition 1	1.0	1.0	1.0	1.0	1.0
Audition 2	1.0	1.0	1.5	2.0	2.0
Audition 3	1.0	1.0	1.0	1.0	1.5
Audition 4	1.0	1.0	1.0	1.0	2.0
Audition 5	1.0	1.0	1.0	1.5	1.5
Audition 6	1.0	1.0	1.0	1.0	1.0
Audition 7	1.0	1.0	1.0	1.0	1.0

Subject 4 | **Sex**: M | **Age**: 35 | **Date**: 16/10/2014

	Round 1 55 dB (SPL)	Round 2 60 dB (SPL)	Round 3 65 dB (SPL)	Round 4 70 dB (SPL)	Round 5 75 dB (SPL)
Audition 1	1.0	1.0	1.0	1.0	1.0
Audition 2	1.0	1.0	1.0	1.0	1.0
Audition 3	1.0	1.0	1.0	1.0	1.0
Audition 4	1.0	1.0	1.0	1.0	1.0
Audition 5	1.0	1.0	1.0	1.0	1.0
Audition 6	1.0	1.0	1.0	1.0	1.0
Audition 7	1.0	1.0	1.0	1.0	1.0

Subject 5 | **Sex**: F | **Age**: 45 | **Date**: 16/10/2014

	Round 1 55 dB (SPL)	Round 2 60 dB (SPL)	Round 3 65 dB (SPL)	Round 4 70 dB (SPL)	Round 5 75 dB (SPL)
Audition 1	1.0	1.0	1.0	1.0	1.0
Audition 2	1.0	1.0	1.0	1.0	1.0
Audition 3	1.0	1.0	1.0	1.0	1.0
Audition 4	1.0	1.0	1.0	1.0	2.0
Audition 5	1.1	1.0	1.0	1.0	1.0
Audition 6	1.0	1.0	1.0	1.0	1.0
Audition 7	1.0	1.0	1.0	1.0	1.0

Subject 6 | **Sex**: M | **Age**: 23 | **Date**: 16/10/2014

	Round 1 55 dB (SPL)	Round 2 60 dB (SPL)	Round 3 65 dB (SPL)	Round 4 70 dB (SPL)	Round 5 75 dB (SPL)
Audition 1	1.0	1.0	1.0	1.0	1.0
Audition 2	1.0	1.0	1.0	1.0	2.0
Audition 3	1.0	1.0	1.0	1.0	1.0
Audition 4	1.0	1.0	2.0	2.0	2.0
Audition 5	1.0	1.0	1.0	2.0	2.0
Audition 6	1.0	1.0	1.0	1.0	1.0
Audition 7	1.0	1.0	1.0	1.0	1.0

Subject 7 | **Sex**: F | **Age**: 23 | **Date**: 16/10/2014

	Round 1 55 dB (SPL)	Round 2 60 dB (SPL)	Round 3 65 dB (SPL)	Round 4 70 dB (SPL)	Round 5 75 dB (SPL)
Audition 1	1.0	1.0	1.0	1.0	1.0
Audition 2	1.0	1.0	1.0	1.0	2.0
Audition 3	1.0	1.0	1.0	2.0	2.0
Audition 4	1.0	1.0	1.0	2.0	2.0
Audition 5	1.0	1.0	1.0	1.0	2.0
Audition 6	1.0	1.0	1.0	1.0	2.0
Audition 7	1.0	1.0	1.0	1.0	1.0

Subject 8 | **Sex**: M | **Age**: 27 | **Date**: 16/10/2014

	Round 1 55 dB (SPL)	Round 2 60 dB (SPL)	Round 3 65 dB (SPL)	Round 4 70 dB (SPL)	Round 5 75 dB (SPL)
Audition 1	1.0	1.0	1.0	1.0	1.5
Audition 2	1.0	1.0	1.0	1.0	1.0
Audition 3	1.0	1.0	1.0	1.0	2.0
Audition 4	1.5	2.0	1.0	1.5	3.0
Audition 5	1.0	1.0	1.0	1.0	1.0
Audition 6	1.0	1.0	1.0	2.0	3.0
Audition 7	1.2	1.0	1.0	1.0	1.0

Subject 9 | **Sex**: M | **Age**: 51 | **Date**: 16/10/2014

	Round 1 55 dB (SPL)	Round 2 60 dB (SPL)	Round 3 65 dB (SPL)	Round 4 70 dB (SPL)	Round 5 75 dB (SPL)
Audition 1	1.0	1.0	1.0	1.0	1.0
Audition 2	1.0	1.0	1.0	1.0	1.0
Audition 3	1.0	1.0	1.0	1.0	1.0
Audition 4	1.0	1.0	1.0	1.0	1.0
Audition 5	1.0	1.0	1.0	1.0	1.0
Audition 6	1.0	1.0	1.0	1.0	1.0
Audition 7	1.0	1.0	1.0	1.0	1.0

Subject 10 | **Sex**: M | **Age**: 26 | **Date**: 16/10/2014

	Round 1 55 dB (SPL)	Round 2 60 dB (SPL)	Round 3 65 dB (SPL)	Round 4 70 dB (SPL)	Round 5 75 dB (SPL)
Audition 1	1.0	1.0	1.0	1.0	1.0
Audition 2	1.0	1.0	1.0	1.0	1.2
Audition 3	1.2	1.3	2.0	2.0	2.1
Audition 4	1.0	1.0	1.3	1.3	1.3
Audition 5	1.0	1.3	2.0	2.0	2.2
Audition 6	1.8	2.0	2.0	2.0	2.8
Audition 7	1.0	1.0	1.0	1.2	1.2

Subject 11 | **Sex**: M | **Age**: 24 | **Date**: 16/10/2014

	Round 1 55 dB (SPL)	Round 2 60 dB (SPL)	Round 3 65 dB (SPL)	Round 4 70 dB (SPL)	Round 5 75 dB (SPL)
Audition 1	1.0	1.0	1.0	1.0	1.0
Audition 2	1.0	1.0	1.0	1.2	1.5
Audition 3	1.2	1.5	1.5	1.5	1.5
Audition 4	1.5	1.5	2.0	2.0	2.0
Audition 5	1.5	1.4	1.5	2.0	2.0
Audition 6	1.5	1.0	1.0	1.0	1.5
Audition 7	1.0	1.5	1.5	1.5	2.0

Subject 12 | **Sex**: F | **Age**: 33 | **Date**: 16/10/2014

	Round 1 55 dB (SPL)	Round 2 60 dB (SPL)	Round 3 65 dB (SPL)	Round 4 70 dB (SPL)	Round 5 75 dB (SPL)
Audition 1	1.0	1.0	1.0	1.0	1.0
Audition 2	1.0	1.0	1.0	1.0	1.0
Audition 3	1.0	1.0	1.0	1.0	1.0
Audition 4	1.0	1.0	1.0	1.0	2.0
Audition 5	1.0	1.0	1.0	1.0	1.0
Audition 6	1.0	1.0	1.0	1.0	1.0
Audition 7	1.0	1.0	1.0	1.0	1.0

Subject 13 | **Sex**: M | **Age**: 17 | **Date**: 23/10/2014

	Round 1 55 dB (SPL)	Round 2 60 dB (SPL)	Round 3 65 dB (SPL)	Round 4 70 dB (SPL)	Round 5 75 dB (SPL)
Audition 1	1.4	1.1	1.0	1.3	1.7
Audition 2	1.1	2.3	2.7	2.7	3.1
Audition 3	1.0	1.0	1.2	1.0	2.3
Audition 4	1.1	1.0	1.0	2.0	3.6
Audition 5	1.9	2.0	1.8	2.6	4.0
Audition 6	1.2	2.3	2.4	1.8	2.0
Audition 7	1.4	1.6	1.4	1.5	2.0

Subject 14 | **Sex**: F | **Age**: 17 | **Date**: 23/10/2014

	Round 1 55 dB (SPL)	Round 2 60 dB (SPL)	Round 3 65 dB (SPL)	Round 4 70 dB (SPL)	Round 5 75 dB (SPL)
Audition 1	1.2	1.0	1.0	1.1	1.1
Audition 2	1.4	1.2	1.0	1.5	1.0
Audition 3	1.0	1.5	2.0	2.0	2.0
Audition 4	2.0	1.5	1.5	2.0	2.0
Audition 5	2.0	1.6	1.6	1.6	2.6
Audition 6	2.1	1.8	1.5	2.4	3.0
Audition 7	1.5	1.6	2.0	1.5	1.5

Subject 15 | **Sex**: F | **Age**: 16 | **Date**: 23/10/2014

	Round 1 55 dB (SPL)	**Round 2** 60 dB (SPL)	**Round 3** 65 dB (SPL)	**Round 4** 70 dB (SPL)	**Round 5** 75 dB (SPL)
Audition 1	1.5	1.1	1.0	1.0	1.0
Audition 2	1.0	1.0	1.0	1.0	1.0
Audition 3	1.0	1.1	1.0	1.0	1.0
Audition 4	1.0	1.0	1.0	1.0	1.2
Audition 5	1.0	1.0	1.0	1.0	1.2
Audition 6	1.1	1.2	1.0	1.0	1.3
Audition 7	1.1	1.1	1.0	1.0	1.0

APPENDIX D

Acoustic Link Budget

A link budget is the accounting of all of the gains and losses from the transmitter, through the medium (free space, cable, waveguide, fiber, etc.), to the receiver in a telecommunication system. It accounts for the attenuation of the transmitted signal due to propagation, the antenna gains, feed-line and other losses. Randomly varying channel gains (such as fading) are considered by adding some margin depending on the anticipated severity of its effects. This margin can be reduced by the use of mitigation techniques such as antenna diversity or frequency hopping. Inspired by the advantages of using a Link Budget analysis in the design process of RF telecommunications links (Figure D.1), we present an analogous methodology for an Acoustic Link Budget. Our aim is to model a complete acoustic ranging system, with transmitter, receiver and propagation channel, in a simple and useful representation.

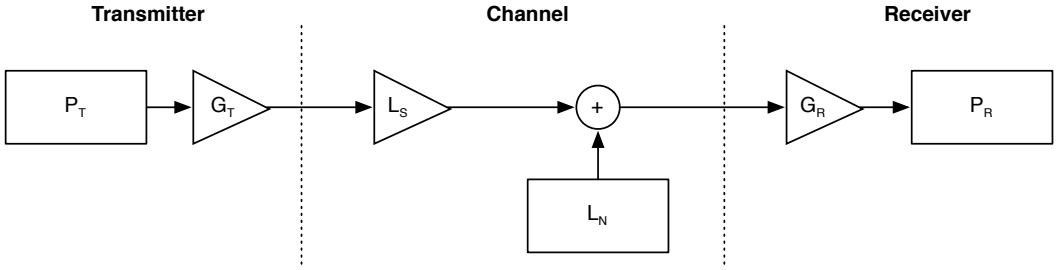


Figure D.1: Typical Link Budget analysis used in telecommunications.

In Figure D.1, P_T defines the transmitted power, G_T is the gain at the transmitter, L_S is the channel losses due to channel attenuation and transducer sensitivity, L_N is the noise factor (i.e. environmental noise, and multi-access interference), G_R is the gain at the receiver and P_R defines the received power.

From this analysis we can obtain the Link Budget Equation, cf. Equation D.1,

$$P_R = P_T + G_T + L_S + L_N + G_R \tag{D.1}$$

Another important figure of merit in the Link Budget Analysis is the Signal-Noise-Ratio (SNR) obtained at the receiver, which is given by Equation D.2.

$$\text{SNR} \quad = \quad P_R - N_F \tag{D.2}$$

From a practical point of view, we have considered that the noise floor N_F figure of merit includes the electrical/thermal and acoustic noise floors at the receiver and can be measured after looking to the received signal in quiescent conditions. With this, we have obtained a noise floor figure of merit based on all noise sources (electrical/thermal + acoustic noise floor), included at the receiver.

Figure D.2 illustrates the Acoustic Link Budget we propose. Transducers were added to include its sensitivity (also known as transducer factor), into the equation. The sound pressure level at the speaker output is proportional to its input voltage, which enabled the use of its sensitivity (S_S) as the transducer factor at the transmitter. Likewise, the output voltage at the microphone output is proportional to the incident sound pressure level, which enabled the use of its sensitivity (S_M) as our transducer factor at the receiver.

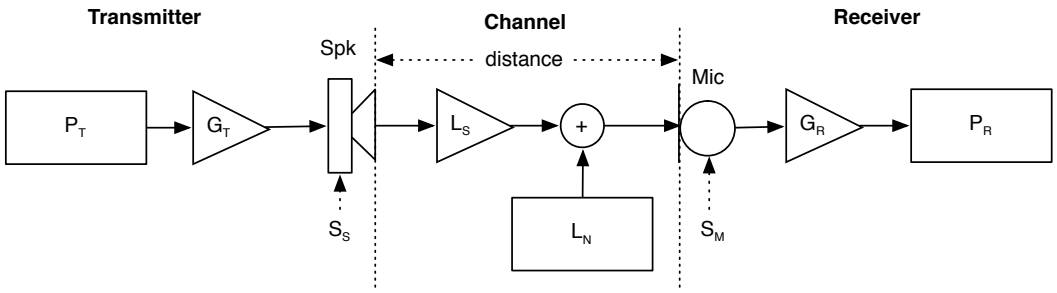


Figure D.2: Acoustic Link Budget.

From this analysis, the Acoustic Link Budget Equation (Equation D.3), is obtained. Note that the Acoustic Link Budget equation must be normalized to the same acoustic pressure reference to avoid errors, and is important to convert the microphone sensitivity (S_M) to dB V/μPa rather than use the standard notation that often appears in dB V/Pa.

$$P_R \quad = \quad P_T + G_T + S_S + L_S + L_N + S_M + G_R \tag{D.3}$$

For example, consider a speaker with sensitivity $S_S = 80$ dB (SPL) (re 20μPa @ 1 m) and a microphone with sensitivity $S_M = -35$ dB (re V/Pa). In this case we must convert the microphone sensitivity to the reference of μPa, cf. Equation D.5.

$$S_{M(V/\mu Pa)} = 20\log_{10}(10^{\frac{-35}{20}} \times 20\ \mu\text{Pa}) \quad (D.4)$$
$$= -129\ dB\ (\text{V}/\mu\text{Pa})$$

To illustrate the procedure on how to use the Acoustic Link Budget to obtain the received power P_R, for a distance of 4 m between a transmitter and a receiver, we apply to the Equation D.3 the following sample data:

- Transmitted Power (P_T) = - 26 dB;
- Transmission Gain (G_T) = 26 dB;
- Speaker Sensitivity (S_S) = 80 dB (SPL) (re 20μPa @ 1 m);
- Channel Losses (L_S) = 6 dB per double distance;
- Microphone Sensitivity (S_M) = -129 dB (re V/μPa), cf. Equation D.5;
- Reception Gain (G_R) = 15 dB;
- Noise Floor at Reception (N_F) = -50 dB.

To perform a link budget analysis for this example we can use the Equation D.3 to obtain the received power (P_R). Furthermore, if the total noise floor can be measured (N_F) we can compute the correspondent Signal-Noise-Ratio (SNR), using Equation D.2. To perform the acoustic link budget analysis, a interactive web-app was developed for easy manipulation of the Acoustic Link Budget Equation (Figure D.3), which can be accessed in the following address `http://www.sergioivanlopes.com/phd/AcousticLinkBudget/`.

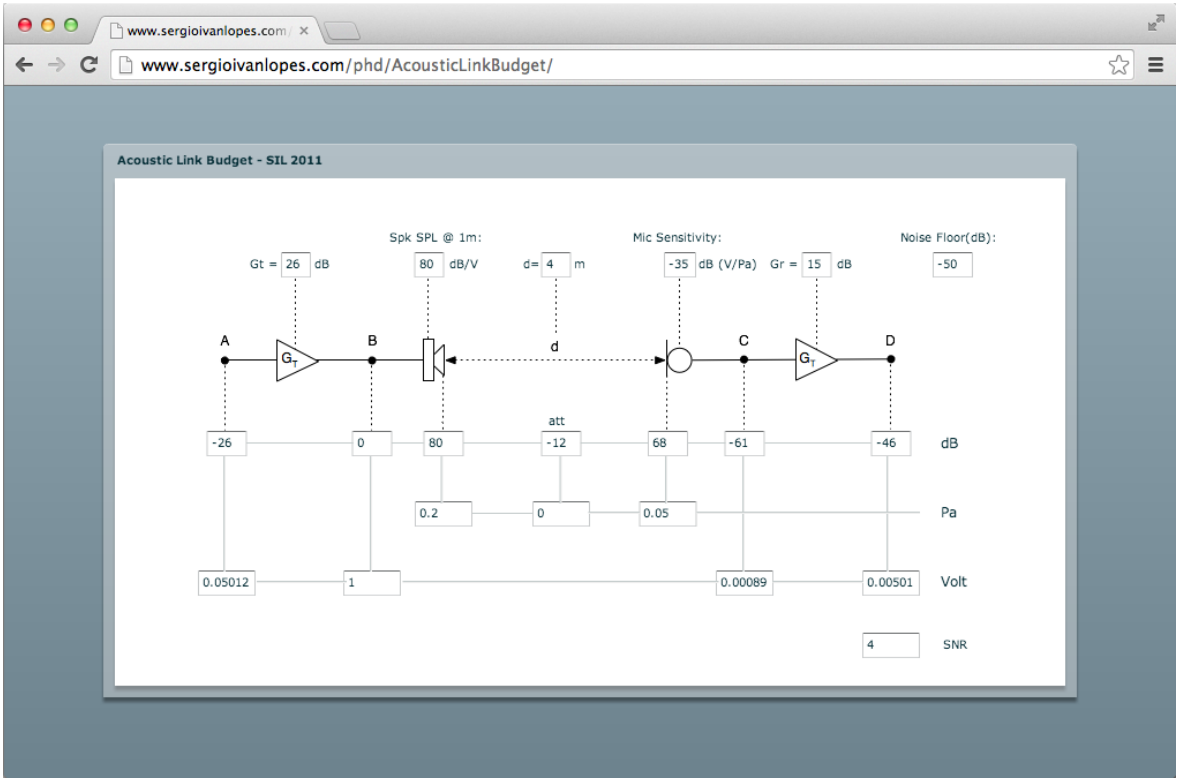


Figure D.3: Acoustic Link Budget interactive web-app.

APPENDIX E

Acoustic Beacon PCB and Schematic

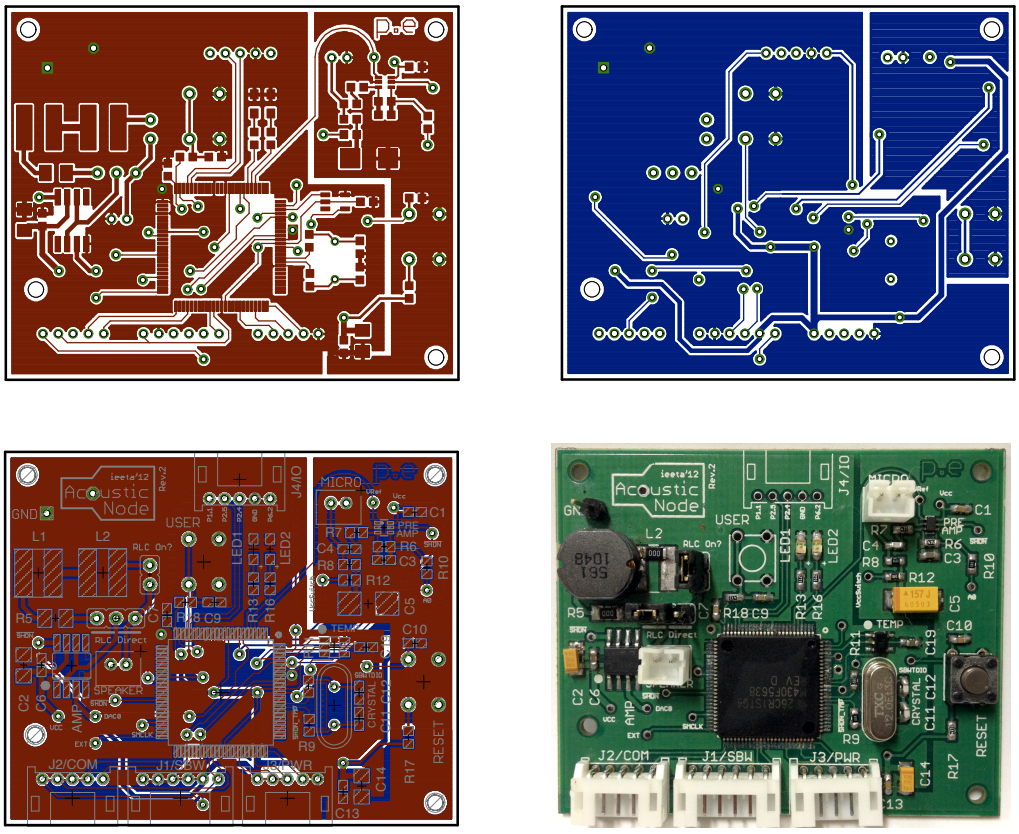


Figure E.1: Acoustic beacon PCB design. Top view, Bottom view, component view and assembled PCB.

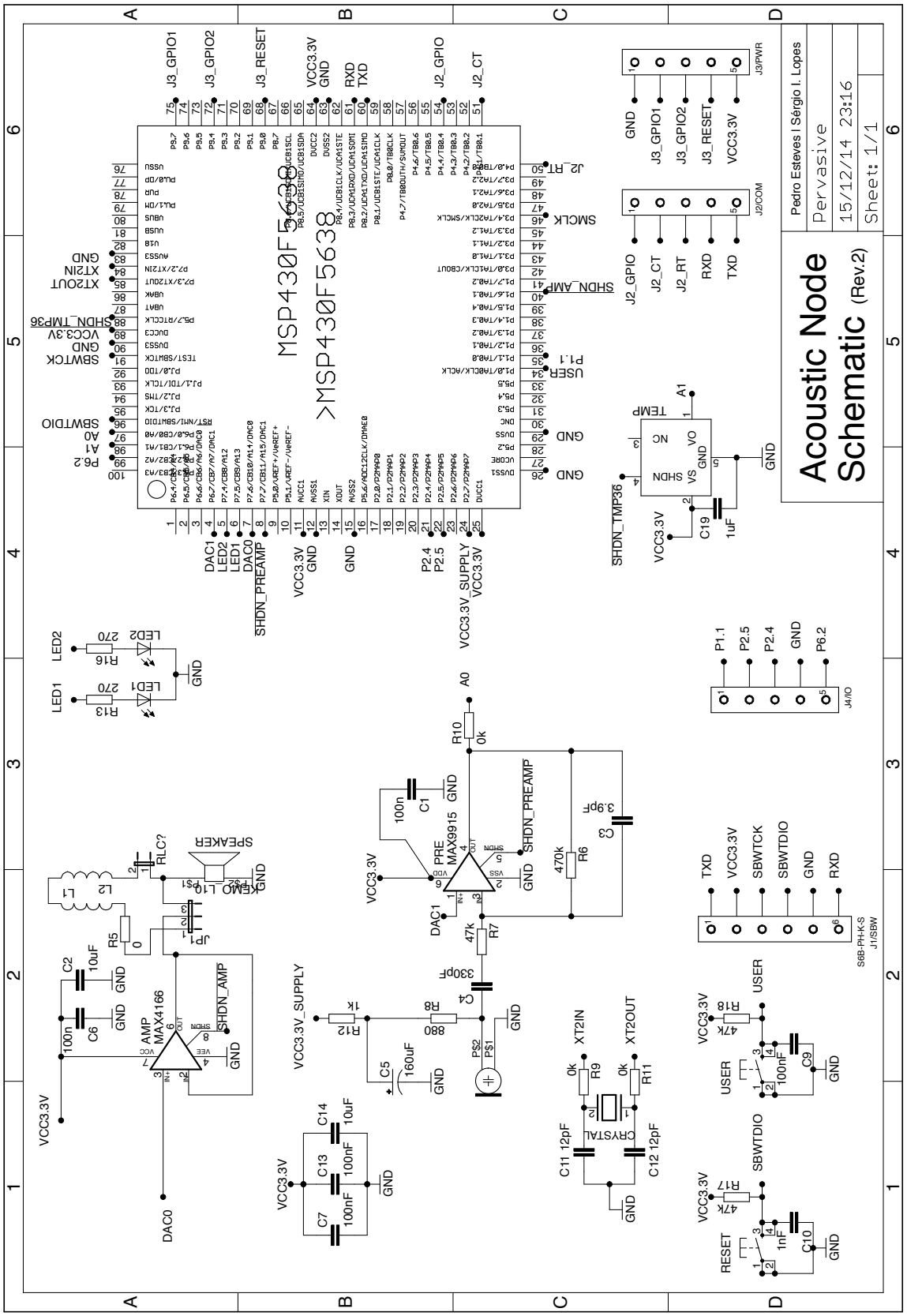


Figure E.2: Acoustic beacon schematic design.